1284
Revised

SEP 24 1973

Agriculture Canada

# Methods for Sensory Evaluation of Food

REVISED

630.4
C212
P 1284
1970
(1973print)
c.2

PUBLICATION 1284

1970

PUBLICATION 1284 1970

# METHODS FOR SENSORY EVALUATION OF FOOD

ELIZABETH LARMOND
Food Research Institute, Central Experimental Farm, Ottawa

CANADA DEPARTMENT OF AGRICULTURE

Copies of this publication may be obtained from
INFORMATION DIVISION
CANADA DEPARTMENT OF AGRICULTURE
OTTAWA
K1A 0C7

©INFORMATION CANADA, OTTAWA, 1973

Printed 1967
Revised 1970
Reprinted 1971, 1973

Code No.: 3M-36513-9:73
Cat. No.: A53-1284

## CONTENTS


Introduction ..... 3

Types of Tests ..... 5

Samples and Their Preparation ..... 5

Panelists ..... 7

Testing Conditions ..... 8

Questionnaires ..... 10

Design of Experiments and Methods of Analyzing Data .... 13

Consumer Testing ..... 14

Appendix I – Sample Questionnaires and Examples of Analyses

- Triangle Test Difference Analysis ..... 15
- Duo-Trio Test Difference Analysis ..... 17
- Multiple Comparison Difference Analysis ..... 19
- Ranking Difference Analysis ..... 24
- Scoring Difference Test ..... 27
- Paired Comparison Difference Test ..... 31
- Hedonic Scale Scoring ..... 36
- Paired Comparison Preference ..... 37
- Ranking Preference ..... 38

Appendix II – Statistical Chart 1 ..... 39

- Statistical Chart 2 – F Distribution ..... 40
- Statistical Chart 3 – Studentized Ranges, 5 percent ... 42
- Statistical Chart 4 – Studentized Ranges, 1 percent ... 43
- Statistical Chart 5 – Scores for Ranked Data ..... 44

Appendix III – Notes on Introductory Statistics by Andres Petrasovits ..... 45

References ..... 55

Additional Sources of Information ..... 56

Digitized by the Internet Archive
in 2012 with funding from
Agriculture and Agri-Food Canada – Agriculture et Agroalimentaire Canada

http://www.archive.org/details/methodsforsensor00larm

## INTRODUCTION

A sensory evaluation is made by the senses of taste, smell, and touch when food is eaten. The complex sensation that results from the interaction of our senses is used to measure food quality in programs for quality control and new product development. This evaluation may be carried out by one person or by several hundred.

The first and simplest form of sensory evaluation is made at the bench by the research worker who develops the new food products. He relies on his own evaluation to determine gross differences in products. Sensory evaluation is conducted in a more formal manner by laboratory and consumer panels.

Most aspects of quality can be measured only by sensory panels, although advances are being made in the development of objective tests that measure individual quality factors. Instruments that measure texture are probably the best known. Some examples are the L.E.E.-Kramer shear press and the Warner-Bratzler shearing device. Gas chromatography and mass spectrometry enable odor to be measured to a limited extent. The color of foods can be accurately measured by tristimulus colorimetry. As new instruments are developed to measure quality, sensory evaluation will be used to prove and standardize new objective tests.

When people are used as a measuring instrument, it is necessary to rigidly control all testing methods and conditions to overcome errors caused by psychological factors. "Error" is not synonymous with mistakes, but may include all kinds of extraneous influences. The physical and mental condition of the panelist and the influence of the testing environment affect sensory tests. For example, some people may have more flavor acuity in the morning, others in the afternoon. Even the weather can influence the disposition of panelists.

Small panels are used to test the palatability of foods. They may also be used in preliminary acceptance testing. Laboratory panels may be used to determine:

the best processing procedures,
suitable varieties of raw material,
preferable cooking and processing temperatures,
effect of substituting one ingredient for another,
best storage procedures,
effect of insecticides and fertilizers on flavor of foods,
effect of animal feeds on flavor and keeping quality of meat,
optimum size of pieces and their importance,
effect of color on acceptability of foods,
suitable recipes for the use of new products,
comparison with competitors' products.

The author is grateful to Mr. A. Petrasovits, Statistical Research Service, Canada Department of Agriculture, who wrote Appendix III. Statistical Chart 1 has been reprinted with permission from Wallerstein Laboratory (Bengtsson, K. 1953. *Wallerstein Lab. Commun.* 16 (No. 54): 231–251.). Thanks are due to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., Cambridge, to Dr. Frank Yates, F.R.S. Rothamstead, and to Messrs. Oliver and Boyd Ltd., Edinburgh, for permission to reprint in part Tables V (Statistical Chart 2) and XX (Statistical Chart 5) from their book *Statistical Tables for Biological, Agricultural and Medical Research.* Thanks are also due to Dr. Malcolm Turner for permission to reprint "Multiple Range and Multiple F Tests" (Statistical Charts 3 and 4) by D. B. Duncan from *Biometrics*, Volume II, 1955.

## TYPES OF TESTS

The two types of tests are difference tests and preference tests.

### Difference Tests

In difference tests the members of the panel are merely asked if a difference exists between two or more samples. Individual likes and dislikes are disregarded and each panelist is advised to be objective in his evaluation. He may not like a particular product, but he should be taught what constitutes good and poor quality and try to evaluate it on the basis of the instructions received. He is acting as a quality measuring instrument.

### Preference Tests

Preference or acceptance tests determine representative population preferences, and need many people on the panel. The total scores from trained panels can be used to predict preference scores obtained from panels of 100 to 160 untrained persons (19). Some tests are conducted on a national scale by firms specializing in this form of testing. Even after extensive tests, there is no assurance that the results will apply to the total population.

## SAMPLES AND THEIR PREPARATION

Panel members are usually influenced by all the characteristics of the test material. Therefore, test samples should be prepared and served as uniformly as possible (6).

### Information About Samples

As little information as possible about the test should be given to the panelists, since this information may influence results. It has been found (13) that if information given to the panelists has meaning within the terms of their experience it will influence their responses; panelists will taste what they expect to taste. When panelists were told that high-quality raw products were used, the panel preference was high, whereas the knowledge that low-quality raw products were used lowered the panel preference. The information that the food had been treated with unspecified chemicals and exposed to unspecified sterilizing rays did not alter panel findings.

### Temperature of Samples

The samples must be of uniform temperature. Therefore the mechanical problems of serving foods at a constant and uniform temperature should be

carefully considered. The temperature at which food is usually eaten is recommended for samples. However, taste buds are less sensitive to very high or very low temperatures, which impair full flavor perception.

### Sample Uniformity

To measure flavor differences in products of large unit size, such as canned peach halves, slice the large units into smaller pieces and carefully mix them to obtain a more uniform sample. To test the quality of canned juice, open several cans and mix all the juice together before preparing individual test samples.

### Coding

Samples should be coded in such a manner that the judges cannot distinguish the samples by the code or be influenced by code bias. For example, if the samples are numbered 1, 2, 3, or lettered A, B, C, a coding bias could be caused because people associate 1 or A with "first" or "best" and might tend to score this sample higher. A set of three-digit random numbers should be assigned to each sample so that the panelists will receive samples coded differently (17).

### Number of Samples

To determine the number of samples to be presented at one testing session (12) consider the following:

*The nature of the product being tested* – No more than six samples of ice cream should be evaluated because of the temperature of the product.

*The itensity and complexity of the sensory property being judged* – It has been found (18) that with mild products such as green beans and canned peaches up to 20 samples may be tasted with no decrease in the taster's ability to discriminate.

*The experience of the taster* – A professional tea, wine, or coffee taster can evaluate hundreds of samples in one day.

*The amount of time and commodity available.*

### Order of Presentation

The order in which the samples are presented to the judges may also influence results. Studies on the effect of sample sequence on food preference (8) have shown position effect (the later samples were rated lower); contrast effect (serving "good" samples first lowered the ratings for "poor" samples); and convergence effect (tendency to make similar responses to successive stimuli). Contrast and convergence effects were shown to be independent of position effect. Random presentation to the panelists equalized these effects.

Figure 1. Preparation of samples.
Figure 2. Presentation of samples through hatch.

## PANELISTS

For economical reasons choose panel members from all available personnel, including office, plant, and research staff. It should be considered a part of work routine for personnel in the food industry to serve as panelists. However, no one should be asked to evaluate foods to which he objects. Persons concerned with the test product (product development and production) and those who prepare the samples for testing should not be included on the panel.

Panel members should be in good health and should excuse themselves if they are suffering from a cold. Nonsmokers and smokers have been found equally useful for panels. It is inadvisable for smokers to smoke within one to two hours before a test (3). Heavy smokers, people who smoke one or more packs of cigarettes per day, have been found to be generally less sensitive than nonsmokers (2). There are exceptions however. No correlation exists between age and sensitivity. A certain amount of sensitivity is required, but this seems not as important as experience (1). A person of average sensitivity, a high degree of personal integrity, ability to concentrate, intellectual curiosity, and willingness to spend time in evaluation may do a better job than a careless person with extreme acuity of taste and smell.

To keep the panelists interested in the work that is being done, show them the results when each series of tests is completed. The importance of the work can be shown by running the tests in a controlled, efficient manner.

### Selection and Training

Panelists should be selected for their ability to detect differences. People who do well with some products often do poorly on others. A taster is seldom equally proficient in tasting all foods.

Researchers disagree on the value of training panelists. It has been stated (17) that screening tests could be used to choose panelists who are capable of detecting differences, and actual training would be unnecessary. At the Operational Research Group of Cadbury Bros. Ltd., in Bournville, England, panelists with sufficiently discriminating palates are selected on the basis of a series of triangular tests with typical confectionery materials. The selected panelists are given 10 weeks' training on special training panels and finally they attend the formal sessions of a tasting panel for 4 weeks before their assessments are actually used (15). Certainly a specific panel for the product and method being tested is more useful than a general-purpose panel. Panelists must be familiar with the product and must know what constitutes good quality in the product. Preliminary sessions will help to clarify the meaning of descriptive terms. To be of any real use terms should be referable to specific objective standards (23).

### Number of Panelists

Since there are many sources of variation in sensory tests, the more tasters on the panel, the more likely it is that the individual variations will balance. Four panelists is probably the minimum number (11), although most researchers think there should be eight or ten (10). A small panel of high sensitivity and ability to differentiate may be preferable to a large panel of less sensitivity.

## TESTING CONDITIONS

### Testing Area

The room where the tests are run may be simple or elaborate, but the panelists must be independent of each other, in separate booths or in partitioned sections at a large table. The panelists must be free from distractions such as noise.

If possible, the room should be odor free and separate from, though adjacent to, the area where the sample is prepared. Air-conditioning is useful in the testing area. The walls should be off-white or a light neutral gray so that sample color will not be altered.

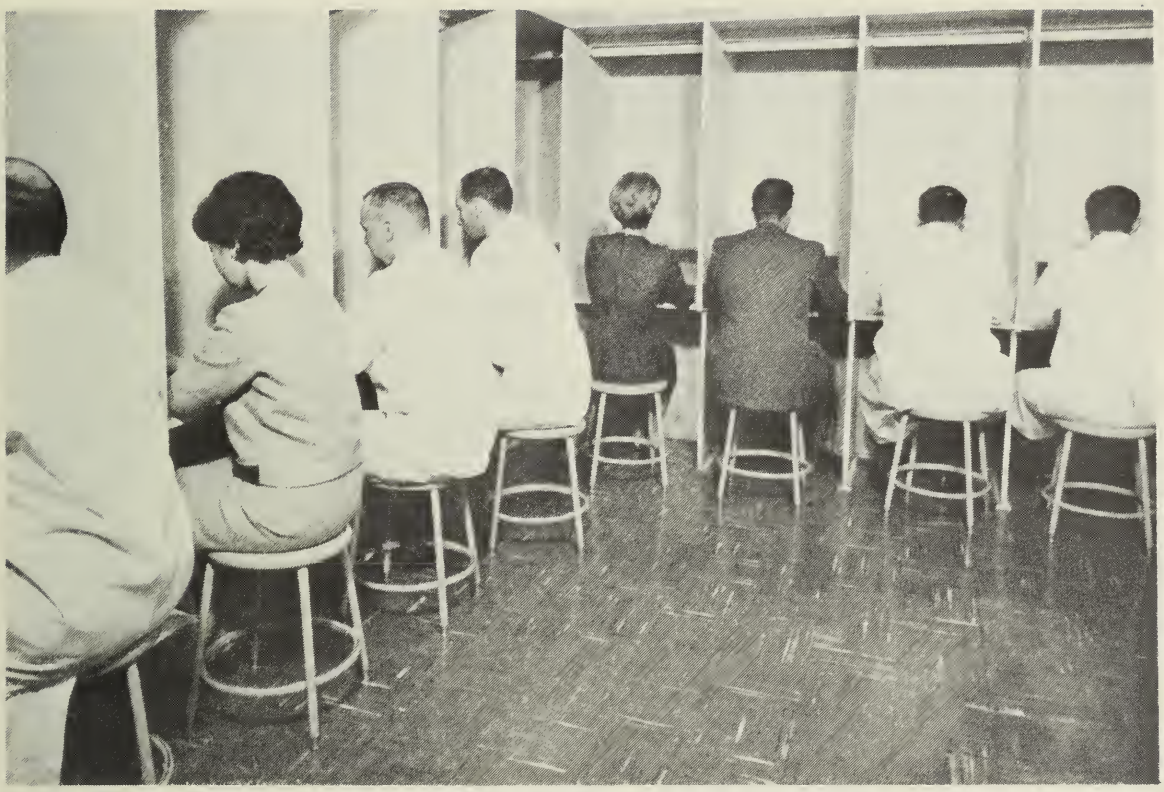


Figure 3. Taste panel area.

### Lighting

Uniform lighting is essential. Natural white fluorescent lights are more suitable than cool white lights. Red fluorescent lights are used in the testing area of the Food Research Institute, Ottawa, to hide obvious color differences in samples whose flavor is being evaluated.

### Testing Schedule

The time of day that tests are run influences results. Although this cannot be controlled if the number of tests is large, late morning and midafternoon have been found to be the best times for testing. Since a panelist's eating habits affect test results, it is desirable that no tests be performed in the period from 1 hour before a meal to 2 hours after a meal.

### Containers

The samples to be tested should be presented to the panelists in clean, odorless, and tasteless containers.

### Tasting Procedures

It is generally agreed that whether a panelist swallows the sample or spits it out the result of the test is unchanged. However, the panelist should be instructed to use the same method with each sample in each test.

Use crackers, white bread, celery, apples, or water to remove all traces of flavor from the mouth between tasting samples of certain foods. If water is used it should be at room temperature, as cold water reduces the efficiency of the taste buds.

## QUESTIONNAIRES

The simplest type of questionnaire has been found to be the most efficient. Elaborate questionnaires divert the attention of the panelist and complicate interpretations. No blanks should be put in the questionnaire except those that apply to the panelist. The person analyzing the test results should use separate summary sheets.

A new questionnaire should be prepared if the method or objective of the test changes. The increased reliability of the test results is worth the time and effort needed to make up new questionnaires.

Questionnaires for some commonly used tests follow. Sample questionnaires with examples of their analyses are shown in Appendix I.

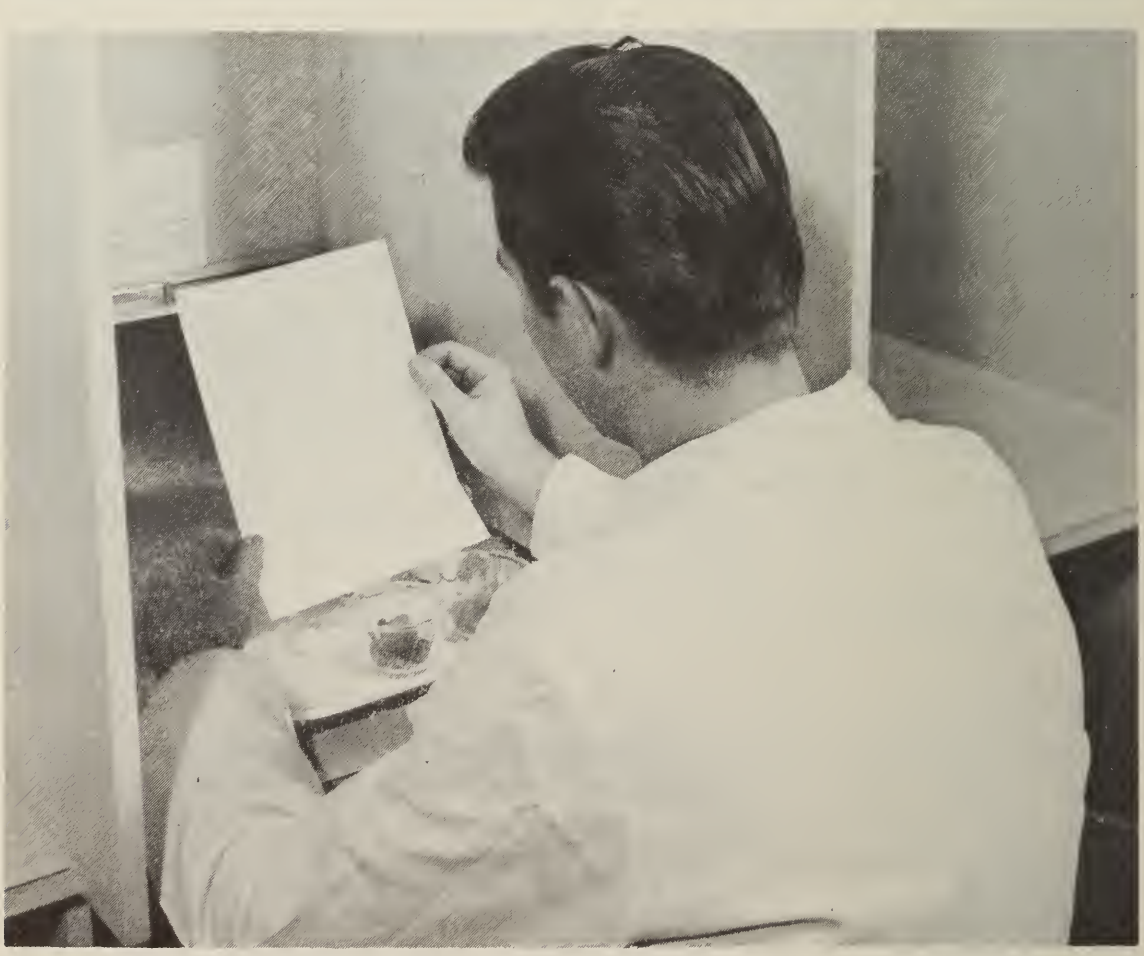
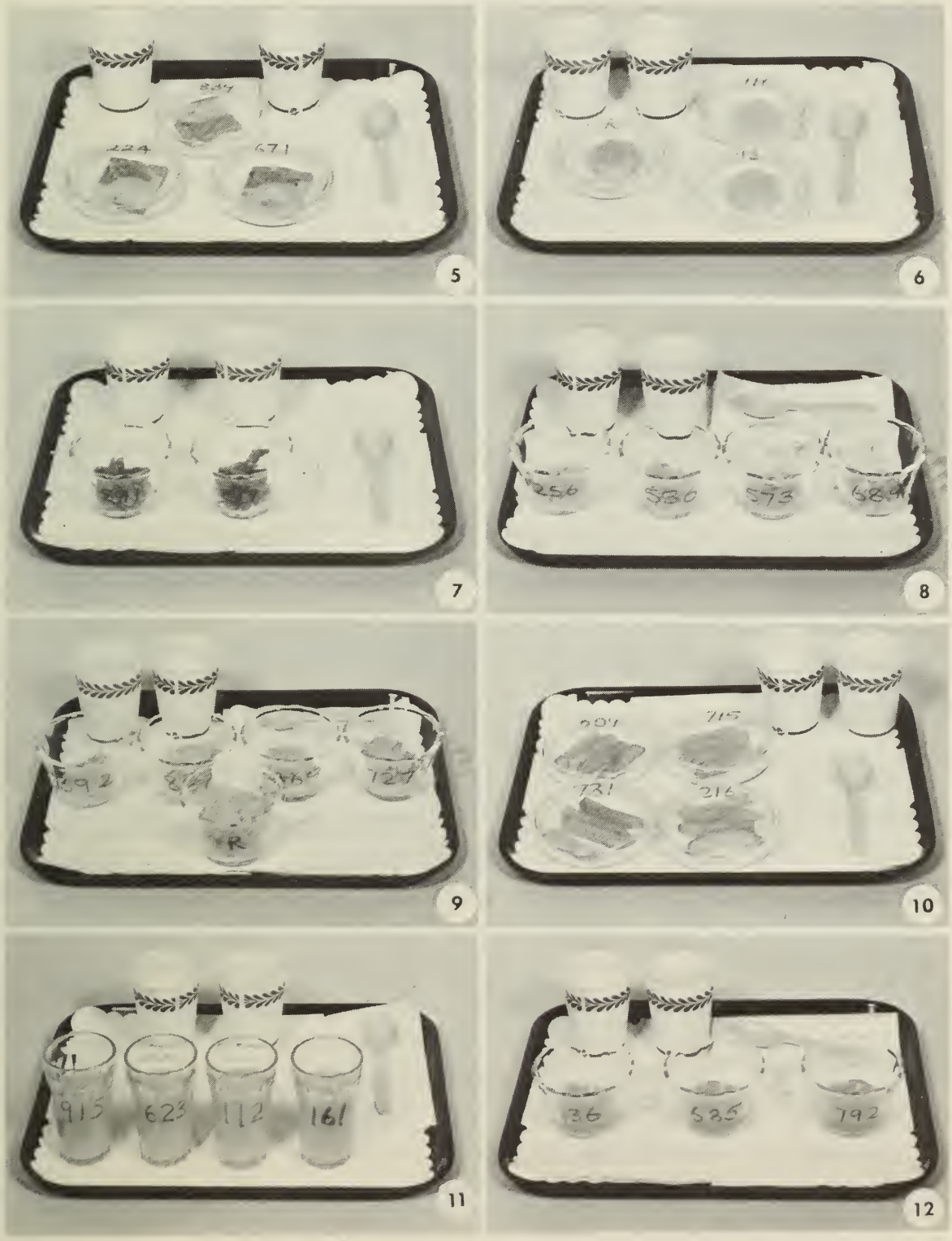


Figure 4. Panelist with samples and questionnaire.




Figures 5–12. Examples of trays prepared for the following tests: 5, triangle test; 6, duo-trio test; 7, paired comparison test; 8, ranking for difference; 9, multiple comparison; 10, scoring for difference; 11, scoring for preference; 12, ranking for preference.

## Difference Tests

*The triangle test* – In the triangle test, three coded samples are presented to the panelist. He is told that two samples are identical and he is asked to indicate the odd one. See sample questionnaire on page 15.

*The duo-trio test* – In the duo-trio test, three samples are presented to the taster, one is labeled R (reference) and the other two are coded. One coded sample is identical with R and the other is different. The panelist is asked to identify the odd sample. See sample questionnaire on page 17.

*Paired comparison test* – In the paired comparison test, a pair of coded samples that represent the standard or control and an experimental treatment are presented to the panelist, who is asked to indicate which sample has the greater or lesser degree of intensity of a specified characteristic – such as sweetness and hardness. If more than two treatments are being considered, each treatment is compared with every other in the series. The design becomes somewhat cumbersome if many treatments are compared.

An example of a paired comparison test is given on page 31.

In most instances the value of a paired comparison can be increased by a panelist's report on the extent of the difference found (20).

Paired testing is generally used to compare new with old procedures and in quality control. When evidence from paired comparison experiments is reported, do not conclude without further facts that a less preferred treatment is of poor quality. The panelist should be asked to rate the quality of each treatment as good, fair, or poor.

*Ranking* – The panelist is asked to rank several coded samples according to the intensity of some particular characteristic. See sample questionnaire on page 24.

*Multiple comparison* – In multiple comparison tests, a known reference or standard sample is labeled R and presented to the panelist with several coded samples. The panelist is asked to score the coded samples in comparison with the reference sample. See sample questionnaire on page 19.

*Scoring* – Coded samples are evaluated by the panelist who records his reactions on a descriptive graduated scale. These scores are given numerical values by the person who analyzes the results. See sample questionnaires on pages 27 and 30.

*The flavor-profile method* – The flavor-profile procedure was developed by Arthur D. Little Incorporated. A small laboratory panel of six or eight people who have been trained in the method measure the flavor profile of food products. Descriptive words and numbers, with identical meaning to each panel member, are used to show the relative strength of each note on a suitable scale. With this method it is possible to determine small degrees

of difference between two samples, the degree of blending, degrees of similarity, and overall impression of the product. Considerable knowledge of flavor is required for the interpretation of flavor-profile results, since they cannot be analyzed statistically. The flavor-profile method requires great skill, extensive education in odor and flavor sensations, and keen interest and intelligence on the part of the panelist. This technique has been reviewed in detail by Sjostrom (21) and Caul (4).

*Dilution tests* – Dilution tests involve the determination of the identification threshold for the material under study. The flavor of a product is described in terms of the percentage of dilution or as a ratio that reflects the actual amount of odor or flavor detected. This method requires suitable standards for comparison and for dilution of the test material and is limited to foods that can be made homogeneous without affecting flavor (24).

### Preference Tests

*Paired comparison* – The paired comparison test used in preference testing is similar to that used for difference testing. When testing preferences, the panelist is asked which sample he prefers and the degree of preference. See sample questionnaire, page 37.

*Scoring* – Many different types of scales have been developed to try to determine a degree of like or dislike for a food. These scales may be worded "excellent," "very good," "good," "poor," or in other similar manner. However, the preference scale that has probably received the most attention in the past 10 years is the nine-point hedonic scale developed at the Quartermaster Food and Container Institute of the United States. Much time and effort was expended to determine which words best express a person's like or dislike of a food. This hedonic scale is the result of these investigations (14). See sample questionnaire on page 36.

*Ranking* – Ranking follows the same procedure as difference testing (p. 12) except that when used as a preference test the questionnaire is worded so that the panelist will indicate his order of preference for the samples. See sample questionnaire on page 38.

## DESIGN OF EXPERIMENTS AND METHODS OF ANALYZING DATA

The accuracy of sensory evaluation tests on food and the reliance that can be placed on their results depend on standardization of testing conditions and use of statistical methods of experimental design and analysis (22).

Plan the experiment in advance so that a simple mathematical model may be applied to the analysis. Many simple mathematical models depend on the independent nature of the data. Experimental design was introduced to

develop this independence and often makes a simple mathematical model applicable. Experimental design makes the test efficient and saves time and material. An excellent reference for experimental design has been written by Cochran and Cox (5). Application of an experimental design should be approved by a statistician.

Random choice of samples and order of presentation help eliminate error. Replication of tests will strengthen the results.

A discussion of the significance of experimental data is usually based on a comparison of what actually happened to what would happen if chance alone were operating. Before applying statistics to the analysis of experimental data, the specific concept of probability should be understood as it applies to particular experimental data. Appendix III explains the concept of probability. The way in which an experiment is conducted usually determines not only whether inferences can be made but also the calculations required to make them.

## CONSUMER TESTING

The consumer test is used to measure consumer acceptance of a product. Although the fate of a food product depends on consumer acceptance, formal studies of consumer preference are recent. Consumer studies are completely separate from laboratory panels, which do not attempt to predict consumer reaction. In this publication methods used in consumer tests are not described. Ideally a consumer test should cover a large sample of the population for whom the product is intended and should involve geographic as well as income sampling. Because of these conditions, consumer tests are often conducted by a specialized company.

## APPENDIX I

### SAMPLE QUESTIONNAIRES AND EXAMPLES OF ANALYSES

### TRIANGLE TEST
### DIFFERENCE ANALYSIS

DATE _____ TASTER _____

PRODUCT _____

Instructions: Here are three samples for evaluation. Two of these samples are duplicates. Separate the odd sample for difference only.

| (1) Sample | (2) Check odd sample |
|---|---|
| 314 | _____ |
| 628 | _____ |
| 542 | _____ |

(3) Indicate the degree of difference between the duplicate samples and the odd sample.

Slight _____ Much _____

Moderate _____ Extreme _____

(4) Acceptability:

Odd sample more acceptable _____

Duplicate samples more acceptable _____

(5) Comments:

**Example:**

To determine if a difference existed between fish-potato flakes processed under two different sets of conditions, a triangle test was used.

The samples were first reconstituted by adding boiling water, then they were coded. On each tray there were three coded samples: two were the same and one was different. Eleven panelists were given two trays each, one after the other, and asked to identify the odd sample on each tray. This made a total of 22 judgments.

The odd sample was correctly identified 19 times. According to Statistical Chart 1 of Appendix II, page 39, 19 correct judgments from 22 panelists in a triangle test are significant at the 0.1 percent level. The conclusion was that a difference existed between the samples. If the number of correct judgments had been less than 12, the conclusion would be that no detectable difference existed between the samples.

The degree of difference indicated by those panelists who correctly chose the odd sample was:

| | | | |
|---|---|---|---|
| Slight | = 1 | Much | = 6 |
| Moderate | = 7 | Extreme | = 5 |

The next part of the triangle test was to choose the more acceptable sample. Of the 19 panelists who correctly identified the odd sample, 14 found the same sample more acceptable. According to Statistical Chart 1, for a two-sample test (there were only two choices at this point), this is below the number required for significance at the 5 percent level. However, the same degree of significance cannot be attached to the results of this secondary question as to the primary question. Once it is determined that a difference exists, another test should be conducted to determine which sample is more acceptable (probably a paired comparison test).

## DUO-TRIO TEST
## DIFFERENCE ANALYSIS

NAME _____

DATE _____

PRODUCT _____

On your tray you have a marked control sample (R) and two coded samples, one is identical with R, the other is different. Which of the coded samples is different from R?

| SAMPLES | CHECK ODD SAMPLE |
|---|---|
| 432 | _____ |
| 701 | _____ |

**Example:**

To determine if methional could be detected when added to Cheddar cheese in amounts of 0.125 ppm and 0.250 ppm a duo-trio test was used. Each tray had a control sample marked R and two coded samples, one with methional added and one control. The duo-trio test was used in preference to the triangle test because less tasting is required to form a judgment using the duo-trio test. This fact becomes important when tasting a substance with a lingering aftertaste, such as methional.

The test was performed on two successive days using eight panelists. Each day the panelists were presented with two trays: one with 0.125 ppm and the other with 0.250 ppm methional added to a coded sample. This made a total of 16 judgments at each level. The results are shown in Table 1.

TABLE 1

| Panelists | Level of methional added, ppm | | | |
|---|---|---|---|---|
| | 1st day | | 2nd day | |
| | 0.125 | 0.250 | 0.125 | 0.250 |
| P1 | X | R | R | R |
| P2 | R | R | R | R |
| P3 | X | R | X | R |
| P4 | R | X | X | R |
| P5 | R | R | R | R |
| P6 | X | R | X | X |
| P7 | R | R | R | R |
| P8 | R | R | R | R |
| Total | 5 | 7 | 5 | 7 |

P = Panelist
X = Wrong
R = Right
0.125 ppm = 10 out of 16 correct
0.250 ppm = 14 out of 16 correct

Consult Statistical Chart 1 of Appendix II, page 39, for 16 panelists in a two-sample test, which shows that 14 correct judgments are significant at the 1 percent level, while 10 are not significant, even at the 5 percent level.

The conclusion is that methional added to Cheddar cheese can be detected at the 0.250 ppm level but not at the 0.125 ppm level.

## MULTIPLE COMPARISON
## DIFFERENCE ANALYSIS

NAME _____ DATE _____

QUESTIONNAIRE:

You are receiving samples of _____ to compare for _____. You have been given a reference sample, marked R, to which you are to compare each sample. Test each sample; show whether it is better than, comparable to, or inferior to the reference. Then mark the amount of difference that exists.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Sample Number | ____ | ____ | ____ | ____ | ____ | ____ | ____ | ____ |
| Better than R | ____ | ____ | ____ | ____ | ____ | ____ | ____ | ____ |
| Equal to R | ____ | ____ | ____ | ____ | ____ | ____ | ____ | ____ |
| Inferior to R | ____ | ____ | ____ | ____ | ____ | ____ | ____ | ____ |

AMOUNT OF DIFFERENCE:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| None | ____ | ____ | ____ | ____ | ____ | ____ | ____ | ____ |
| Slight | ____ | ____ | ____ | ____ | ____ | ____ | ____ | ____ |
| Moderate | ____ | ____ | ____ | ____ | ____ | ____ | ____ | ____ |
| Much | ____ | ____ | ____ | ____ | ____ | ____ | ____ | ____ |
| Extreme | ____ | ____ | ____ | ____ | ____ | ____ | ____ | ____ |

COMMENTS:

Any comments you may have about the flavor of the samples may be made here:

**Example:**

A multiple comparison test was conducted to determine how much anti-oxidant could be added to fish-potato flakes without tasters detecting a

difference in flavor. The flakes tested contained no antioxidant (0), 1 unit, 2 units, 4 units, and 6 units of antioxidant. Each tray contained a reference sample labeled R that contained no antioxidant and five coded samples (the four different levels of antioxidant and one sample with no antioxidant). Fifteen panelists were asked to evaluate these samples according to the score sheets on page 19. The ratings were given numerical values 1 to 9 by the person analyzing the results with "no difference" equaling 5, "extremely better than R" equaling 1, and "extremely inferior to R" equaling 9. The analysis of variance was calculated as shown in Table 2 and on the following pages.

**TABLE 2**

| Panelists | Level of antioxidant added | | | | | Total |
|---|---|---|---|---|---|---|
| | 0 | 1 Unit | 2 Units | 4 Units | 6 Units | |
| P1 | 1 | 4 | 5 | 1 | 9 | 20 |
| P2 | 3 | 3 | 5 | 5 | 7 | 23 |
| P3 | 7 | 3 | 4 | 4 | 7 | 25 |
| P4 | 5 | 7 | 7 | 3 | 9 | 31 |
| P5 | 3 | 3 | 3 | 3 | 1 | 13 |
| P6 | 1 | 1 | 1 | 1 | 2 | 6 |
| P7 | 5 | 5 | 3 | 5 | 6 | 24 |
| P8 | 2 | 2 | 3 | 2 | 5 | 14 |
| P9 | 1 | 3 | 3 | 3 | 3 | 13 |
| P10 | 1 | 1 | 1 | 7 | 5 | 15 |
| P11 | 6 | 5 | 1 | 4 | 1 | 17 |
| P12 | 7 | 2 | 1 | 3 | 9 | 22 |
| P13 | 3 | 2 | 3 | 2 | 6 | 16 |
| P14 | 3 | 3 | 1 | 5 | 1 | 13 |
| P15 | 3 | 1 | 5 | 3 | 3 | 15 |
| Total | 51 | 45 | 46 | 51 | 74 | 267 |

P = Panelist

**Analysis of Variance**

Correction factor = (Total)$^2$/Number of responses (15 tasters × 5 samples)

$$CF = (267)^2/75 = 71289/75 = 950.52$$

Sum of squares, samples = (Sum of the square of the total for each sample/ Number of judgments for each sample) – CF

$$= [(51^2 + 45^2 + 46^2 + 51^2 + 74^2)/15] - CF$$

$$= (14819/15) - CF = 987.93 - 950.52$$

$$= 37.41$$

Sum of squares, panelists = (Sum of the square of the total for each panelist/Number of judgments by each panelist) – CF

$= [(20^2 + 23^2 + 25^2 \ldots + 15^2)/5] - CF$

$= (5309/5) - CF = 1061.80 - 950.52$

$= 111.28$

Total sum of squares = Sum of the square of each judgment – CF

$= (1^2 + 3^2 + 7^2 + \ldots + 3^2) - CF$

$= 1301.00 - 950.52$

$= 350.48$

The analysis of variance chart was then set up as follows:

| Source of variance | df | SS | MS | F |
|---|---|---|---|---|
| Samples | 4 | 37.41 | 9.35 | 2.59* |
| Panelists | 14 | 111.28 | 7.95 | 2.21* |
| Error | 56 | 201.81 | 3.60 | |
| Total | 74 | 350.48 | | |

df – Degrees of freedom for samples is the number of samples minus one. There were five samples, so the degrees of freedom for samples in this example is 4. Degrees of freedom for panelists is the number of panelists minus one. There were 15 panelists so the df for this source is 14. The df for total is the total number of judgments minus 1 (75 – 1 = 74).

Error – (1) To determine the df for "error" subtract the values obtained for the other variables (in this case 4 for samples and 14 for panelists) from the total, 74,
i.e. 74 – (4 + 14) = 56.

(2) To determine the SS for "error" subtract the values obtained for the other variables (in this case 37.41 for samples and 111.28 for panelists) from the total, 350.48,
i.e. 350.48 – (37.41 + 111.28) = 201.81.

MS – The mean square for any variable is determined by dividing the SS by its respective degree of freedom.

F – The variance ratio or F value for samples is determined by dividing the MS for samples by the MS for error or 9.35/3.60 = 2.59. The F value for panelists may be determined by dividing the MS for panelists by the MS for error.

To determine if the difference between the samples is significant, the calculated F value (2.59) is checked in Chart 2 of Appendix II on pages 40 and 41. With 4 degrees of freedom in the numerator and 56 degrees of freedom in the denominator, the variance ratio (F value) must exceed 2.52 to be significant at the 5 percent level and it must exceed 3.65 to be significant at the 1 percent level. The value of 2.59 is therefore significant at the 5 percent level (*). If the variance ratio is not significant, the conclusion is that the addition of up to 6 units of antioxidant does not make a detectable difference in the flavor of antioxidant.

Since there is a significant difference between the samples, the ones that are different can be determined by using Duncan's Multiple Range Test (7, 16).

| | 0 | 1 Unit | 2 Units | 4 Units | 6 Units |
|---|---|---|---|---|---|
| Sample score = | 51 | 45 | 46 | 51 | 74 |
| Sample mean = Score/Number of panelists = | 51/15 | 45/15 | 46/15 | 51/15 | 74/15 |
| | = 3.4 | 3.0 | 3.1 | 3.4 | 4.9 |

The sample means are arranged according to magnitude:

| A | B | C | D | E |
|---|---|---|---|---|
| 6 Units | 4 Units | 0 | 2 Units | 1 Unit |
| 4.9 | 3.4 | 3.4 | 3.1 | 3.0 |

The standard error of the sample mean:

$$SE = \sqrt{}\ (\text{MS error/Number of judgments for each sample})$$
$$= \sqrt{}\ (3.60/15)$$
$$= \sqrt{}\ 0.24 \qquad = 0.49$$

The "shortest significant ranges" for 2, 3, 4, and 5 means are determined by using Chart 3 of Appendix II, on page 42, for the 5 percent level of probability and Chart 4 for the 1 percent level. In this case to determine the 5 percent level, Chart 3 is used to find the "Studentized ranges" rp for p = 2, . . . 5 means with 56 degrees of freedom (since 56 is not shown, the figure 60 is used).

These values are then multiplied by the standard error of the mean, to obtain the shortest significant ranges, Rp. This gives:

| P | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| rp (5 percent) | 2.83 | 2.98 | 3.08 | 3.14 |
| Rp | 1.39 | 1.46 | 1.51 | 1.54 |

The differences between the sample means are compared with the shortest significant range appropriate for the range under consideration in the following order:

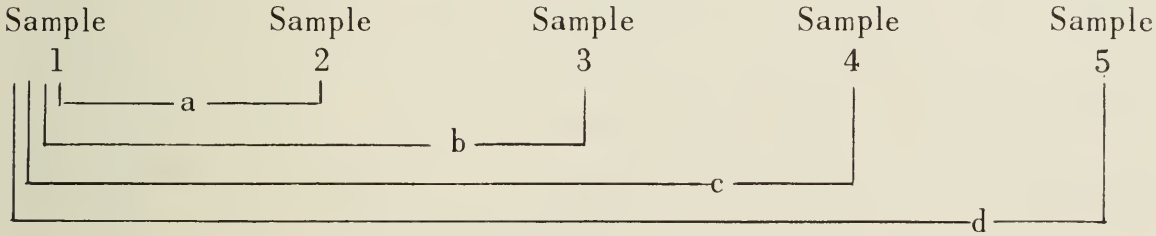
(i) highest minus the lowest, highest minus second lowest, and so on to highest minus second highest;

(ii) second highest minus lowest and so on to second highest minus third highest;

(iii) and so on down to second lowest minus lowest.

If, at any stage, a difference in (i) does not exceed the shortest significant range, the procedure stops, Say, for example, the highest mean minus the *kth* mean does not exceed the shortest significant range, then a line is drawn underscoring all means between the highest and the *kth* means inclusive, which indicates that this set of means should be grouped together as exhibiting no significant differences.

Determination of shortest significant ranges




Range a = 2

b = 3

c = 4

d = 5

(i)

$A - E = 4.9 - 3.0 = 1.9 > 1.54 \quad (R_5)$

$A - D = 4.9 - 3.1 = 1.8 > 1.51 \quad (R_4)$

$A - C = 4.9 - 3.4 = 1.5 > 1.46 \quad (R_3)$

$A - B = 4.9 - 3.4 = 1.5 > 1.39 \quad (R_2)$

A is underscored as it is significantly different from the others.

A B C D E

(ii)

$B - E = 3.4 - 3.0 = 0.4 < 1.51 \quad (R_4)$

Samples B, C, D, and E are underscored together as they exhibit no significant differences.

Proceed no further, since there is no significant difference between B and E.

Therefore, the conclusion is that at the 5 percent level A is significantly different from E or that 6 units made a significant difference in flavor from 0. This procedure may be repeated using Chart 4 if the samples were significantly different at the 1 percent level.

This procedure can be repeated to see which panelists are significantly different from each other.

## RANKING

## DIFFERENCE ANALYSIS

NAME _____

DATE _____

PRODUCT _____

Evaluate these samples for tenderness. Please *rank* the samples for tenderness. The sample that is tenderest is ranked first, the second tenderest is ranked second, the toughest sample is ranked third. Place the code number in the appropriate box.

| 1 | 2 | 3 |
|---|---|---|
| ☐ | ☐ | ☐ |

**Example:**

A ranking test was used to compare the texture of meat of three breeds of geese. The cooked meat was cut into pieces ½ inch by ½ inch by 1 inch and one coded sample from each breed was presented to each of eight panelists. These panelists ranked the samples according to the score sheet above.

Results:

| Ranks: | $B_1$ | $B_2$ | $B_3$ |
|---|---|---|---|
| P1 | 2 | 1 | 3 |
| P2 | 2 | 1 | 3 |
| P3 | 2 | 1 | 3 |
| P4 | 1 | 2 | 3 |
| P5 | 1 | 3 | 2 |
| P6 | 2 | 1 | 3 |
| P7 | 2 | 1 | 3 |
| P8 | 1 | 2 | 3 |
| Total | 13 | 12 | 23 |

P = Panelist
B = Breed
1 = First
2 = Second
3 = Third

To analyze these results the ranks were transformed into scores, according to Fisher and Yates (9). Chart 5 of Appendix II, page 44, was used to determine the numerical value for each score. The sample ranked first of three samples was given a value of 0.85. When converting ranks, the middle rank is given a value of zero and the ranks beyond the middle are given negative values corresponding to the positive values given in the chart. In this case second is 0 and third is −0.85. If we had six ranks the values would be:

first = 1.27
second = 0.64
third = 0.20
fourth = −0.20
fifth = −0.64
sixth = −1.27

In Chart 5 values can be assigned to ranks in tests with up to 30 samples in the manner described above.

| Scores | $B_H$ | $B_P$ | $B_C$ | Total |
|---|---|---|---|---|
| P1 | 0 | 0.85 | −0.85 | 0 |
| P2 | 0 | 0.85 | −0.85 | 0 |
| P3 | 0 | 0.85 | −0.85 | 0 |
| P4 | 0.85 | 0 | −0.85 | 0 |
| P5 | 0.85 | −0.85 | 0 | 0 |
| P6 | 0 | 0.85 | −0.85 | 0 |
| P7 | 0 | 0.85 | −0.85 | 0 |
| P8 | 0.85 | 0 | −0.85 | 0 |
| Total | 2.55 | 3.40 | −5.95 | |

The scores were then analyzed by the analysis of variance, as on page 20.

$$CF = 0$$

$$SS\ samples = (\boxed{2}.55^2 + 3.40^2 + (-5.95)\boxed{2}/8) - CF$$

$$= (53.465/8) - 0$$

$$= 6.68$$

$$SS\ panelists = 0/3 = 0$$

$$Total\ SS = [0^2 + 0^2 + 0^2 + 0.85^2 \ldots + (-0.85)^2] - CF$$

$$= 11.56$$

| Variables | df | SS | MS | F |
|---|---|---|---|---|
| Samples | 2 | 6.68 | 3.34 | 9.54** |
| Panelists | 7 | 0 | | |
| Error | 14 | 4.88 | 0.35 | |
| Total | 23 | 11.56 | | |

| Samples | $B_H$ | $B_P$ | $B_C$ |
|---|---|---|---|
| | 2.55 | 3.40 | −5.95 |
| Mean samples | 0.32 | 0.43 | −0.74 |

| A | B | C |
|---|---|---|
| +0.43 | 0.32 | −0.74 |

$$Standard\ error = \sqrt{(0.35/8)} = \sqrt{0.04375}$$

$$= 0.209$$

| | 2 | 3 |
|---|---|---|
| rp (5 percent) | 3.03 | 3.18 |
| Rp | 0.63 | 0.66 |

$$A - C = 0.43 - (-0.74) = 1.17 > 0.66\ (R_3)$$

$$A - B = 0.43 - 0.32 = 0.11 < 0.63\ (R_2)$$

A B C

$$B - C = 0.32 - (-0.74) = 1.06 > 0.63\ (R_2)$$

C is significantly different from A and B.

The conclusion is that the meat from breed C was significantly less tender than that of breeds H and P at the 5 percent level.

SCORING

DIFFERENCE TEST

NAME _____

DATE _____

PRODUCT _____

Evaluate these samples for flavor. Taste test each one. Use the appropriate scale to show your evaluation and check the point that best describes your feeling about the flavor of the sample.

| Code 815 | Code 558 | Code 394 |
|---|---|---|
| _____ Excellent | Excellent | Excellent |
| _____ Very good | Very good | Very good |
| _____ Good | Good | Good |
| _____ Fair | Fair | Fair |
| _____ Poor | Poor | Poor |
| _____ Very poor | Very poor | Very poor |
| Reason | Reason | Reason |
| | | |

**Example:**

Taste tests were conducted to determine any difference in flavor in the meat of three breeds of geese. The scoring test was used rather than multiple comparison as there was no standard or reference with which to compare. Eight panelists rated the coded samples according to the score sheet on page 27. The ratings were given numerical values by the person analyzing results with excellent = 1, very poor = 6.

The results were analyzed by the analysis of variance, see page 20.

| | $B_1$ | $B_2$ | $B_3$ | Total |
|---|---|---|---|---|
| P1 | 3 | 2 | 3 | 8 |
| P2 | 4 | 6 | 4 | 14 |
| P3 | 3 | 2 | 3 | 8 |
| P4 | 1 | 4 | 2 | 7 |
| P5 | 2 | 4 | 2 | 8 |
| P6 | 1 | 3 | 3 | 7 |
| P7 | 2 | 6 | 4 | 12 |
| P8 | 2 | 6 | 2 | 10 |
| Total | 18 | 33 | 23 | 74 |

Correction factor $= 74^2/24 = 5476/24 = 228.17$

SS samples $= (1/8)\,(18^2 + 33^2 + 23^2) - CF$

$= (1/8) \times 1942 - CF$

$= 242.75 - 228.17$

$= 14.58$

SS panelists $= (1/3)\,(8^2 + 14^2 + \ldots + 10^2) - CF$

$= (1/3) \times 730 - CF$

$= 243.33 - 228.17$

$= 15.16$

Total SS $= (3^2 + 4^2 + \ldots + 2^2) - CF$

$= 276 - CF$

$= 276 - 228.17$

$= 47.83$

| Variables | df | SS | MS | F |
|---|---|---|---|---|
| Samples | 2 | 14.58 | 7.29 | 5.65* |
| Panelists | 7 | 15.16 | 2.17 | |
| Error | 14 | 18.09 | 1.29 | |
| Total | 23 | 47.83 | | |

There is a significant difference between samples at the 5 percent level.

The Multiple Range Test is used to determine which samples are significantly different from the others.

| | $B_1$ | $B_2$ | $B_3$ |
|---|---|---|---|
| Mean samples | = 18/8 | 33/8 | 23/8 |
| | = 2.25 | 4.13 | 2.88 |
| Ranked means | = A | B | C |
| | $B_2$ | $B_3$ | $B_1$ |
| | 4.13 | 2.88 | 2.25 |

Standard error $= \sqrt{(1.29/8)} = \sqrt{0.16}$

$= 0.4$

| P | 2 | 3 |
|---|---|---|
| rp (5 percent) | 3.03 | 3.18 |
| Rp | 1.21 | 1.27 |

$A - C = 4.13 - 2.25 = 1.88 > 1.27\ (R_3)$

$A - B = 4.13 - 2.88 = 1.25 > 1.21\ (R_2)$

A is significantly different from B and C.

$B - C = 2.88 - 2.25 = 0.63 < 1.21\ (R_2)$

A B C

B and C are not significantly different from each other. The conclusion is that breed 2 is significantly different from breeds 1 and 3 at the 5 percent level.

## SCORING

## DIFFERENCE TEST

NAME _____

DATE _____

Evaluate these samples of goose meat for tenderness. Taste test each one. Use the appropriate scale to show your evaluation by checking at the point that best describes your feeling about the sample.

| Code 664 | Code 758 | Code 708 |
|---|---|---|
| _____ Extremely tender | Extremely tender | Extremely tender |
| _____ Very tender | Very tender | Very tender |
| _____ Moderately tender | Moderately tender | Moderately tender |
| _____ Slightly tender | Slightly tender | Slightly tender |
| _____ Slightly tough | Slightly tough | Slightly tough |
| _____ Moderately tough | Moderately tough | Moderately tough |
| _____ Very tough | Very tough | Very tough |
| _____ Extremely tough | Extremely tough | Extremely tough |
| Reason | Reason | Reason |

Note: This is another example of scoring for difference. Analysis of variance is used to analyze results (see page 20).

Numerical values: extremely tender = 1
extremely tough = 8

## PAIRED COMPARISON
## DIFFERENCE TEST

DATE _____ TASTER _____

PRODUCT _____

Evaluate these two samples of peaches for texture.

1. Is there a difference in texture between the two samples?

   Yes _____ No _____

2. Indicate the degree of difference in *texture* between the two samples by checking one of the following statements.

   846 is extremely better than 165
   846 is much better than 165
   846 is slightly better than 165

   No difference

   165 is slightly better than 846
   165 is much better than 846
   165 is extremely better than 846

3. Rate the texture of the samples.

| 165 | 846 |
|---|---|
| Good _____ | Good _____ |
| Fair _____ | Fair _____ |
| Poor _____ | Poor _____ |

Comments:

One of the tests conducted as part of a study of the effect of small doses of irradiation on the keeping quality of fresh peaches was a paired comparison test on texture.

Four samples were compared:
(1) control or 0 krads sample, (2) 150 krads sample, (3) 200 krads sample, and (4) 250 krads sample.

Each sample was compared with every other sample, making a total of six pairs. Each pair was presented to eight panelists for evaluation according to the score sheet on page 31. The experimental design requires that half the panelists taste one sample of the pair first and that the others taste the second sample first.

The ratings of the panelists were given numerical values of +3, +2, +1, 0, −1, −2, −3. Example:

| | Sample | Code |
|---|---|---|
| 1 pair = | 150 krads | 846 |
| | 250 krads | 165 |

The score sheet for the four panelists tasting sample 846 first was set up as follows with the values on the right being assigned by the analyzer.

| | |
|---|---|
| 846 is extremely better than | 165 (+3) |
| 846 is much better than | 165 (+2) |
| 846 is slightly better than | 165 (+1) |
| No difference | ( 0) |
| 165 is slightly better than | 846 (-1) |
| 165 is much better than | 846 (-2) |
| 165 is extremely better than | 846 (-3) |

The four panelists who tasted sample 165 first received a score sheet set up as follows with the corresponding values on the right.

| | |
|---|---|
| 165 is extremely better than | 846 (+3) |
| 165 is much better than | 846 (+2) |
| 165 is slightly better than | 846 (+1) |
| No difference. | ( 0) |
| 846 is slightly better than | 165 (-1) |
| 846 is much better than | 165 (-2) |
| 846 is extremely better than | 165 (-3) |

If one of the four panelists tasting 165 first indicated that 846 was much better than 165, his score was -2.

The results of the scoring for all six pairs by all tasters was tabulated as shown in Table 3.

TABLE 3

| Order of presentation | Frequency of scores equal to | | | | | | | Total score | Mean | Average preference |
|---|---|---|---|---|---|---|---|---|---|---|
| | -3 | -2 | -1 | 0 | +1 | +2 | +3 | | | |
| 0,150 | | | 2 | | 1 | 1 | | 1 | 0.25 | 1.00 |
| 150,0 | | 3 | 1 | | | | | -7 | -1.75 | |
| 0,200 | | | | 1 | 1 | 2 | | 5 | 1.25 | 0.75 |
| 200,0 | | 2 | | | 1 | 1 | | -1 | -0.25 | |
| 0,250 | | | | | 1 | 2 | 1 | 8 | 2.00 | 1.25 |
| 250,0 | | 2 | | | 2 | | | -2 | -0.50 | |
| 150,200 | | 1 | 1 | | 2 | | | -1 | -0.25 | -0.25 |
| 200,150 | | | 1 | 1 | 2 | | | 1 | 0.25 | |
| 150,250 | | | | 2 | 1 | 1 | | 3 | 0.75 | 0.625 |
| 250,150 | | | 3 | | 1 | | | -2 | -0.50 | |
| 200,250 | | 1 | 1 | 2 | | | | -3 | -0.75 | -0.75 |
| 250,200 | | | | 2 | 1 | 1 | | 3 | 0.75 | |
| Total | 0 | 9 | 9 | 8 | 13 | 8 | 1 | | | |

Mean = Total score/Number of panelists = 1/4 = 0.25

Average preferences = ½ (mean for 0,150 - mean for 150,0)

= ½ (0.25-(-1.75)) = ½ (2.00) = 1.00

The average preference of 0 over 150 was 1.00 and the average preference of 150 over 0 was -1.00. Thus the average preference of 0 over 200 equals - the average preference of 200 over 0.

The above data were used to perform an analysis of variance, according to the method of Scheffé (20).

Main effects of treatments ($\hat{\alpha}$) were calculated by totaling the average preference of each sample over every other sample and dividing by the number of treatments.

$\hat{\alpha}_0$ = ¼ (average preference of 0 over 150 + average preference of 0 over 200 + average preference of 0 over 250)

= ¼ (1.00 + 0.75 + 1.25)

= ¼ (3.00) = 0.75

$\hat{\alpha}_{150} = \frac{1}{4}(-1.00 - 0.25 + 0.625) = -0.15625$

$\hat{\alpha}_{200} = \frac{1}{4}(-0.75 + 0.25 - 0.75) = -0.3125$

$\hat{\alpha}_{250} = \frac{1}{4}(-1.25 - 0.625 + 0.75) = -0.28125$

The order effect ($\hat{\delta}$) was calculated by totaling the mean for each order of each pair and dividing this total by the number of ordered pairs (6 pairs × 2 orders = 12).

$\hat{\delta} = (1/12)(0.25 - 1.75 + 1.25 - 0.25 + 2.00 - 0.50 - 0.25 + 0.25 + 0.75 - 0.50 - 0.75 + 0.75)$

$= 0.104$

**ANALYSIS OF VARIANCE TABLE**

| Variables | df | SS | MS | F |
|---|---|---|---|---|
| Main effects | 3 | 24.4375 | 8.1458 | 4.84** |
| Order effect | 1 | 0.5192 | 0.5192 | |
| Error | 44 | 74.0433 | 1.6828 | |
| Total | 48 | 99.0000 | | |

Sum of squares for main effects = number of panelists × number of treatments × sum of squares of each $\hat{\alpha}$

$= 8 \times 4 \times [0.75^2 + (-0.15625)^2 + (-0.3125)^2 + (-0.28125)^2] = 24.4375$

Sum of squares for order effect = number of panelists × number of pairs × order effect $^2(\hat{\delta}^2)$

$= 8 \times 6 \times 0.104^2 = 0.5192$

Total sum of squares (using the frequency of each score as shown in Table 3)

$= 3^2(0 + 1) + 2^2(9 + 8) + 1^2(9 + 13) + 0^2(8)$

$= 99.00$

Sum of squares for error = 99.00 − 24.4375 − 0.5192 = 74.0433

df-*Main effects*. There were 4 treatments, so the degrees of freedom = 3.

*Order effect*. There were 2 orders — one sample of pair first or the other sample of the pair first, df = 1.

*Total*. For paired comparison, the df for total is the total number of observations = 48.

*Error*. 48 − 3 − 1 = 44.

MS for main effects = 24.4375/3 = 8.1458

MS for order effect = 0.5192/1 = 0.5192

MS for error = 74.0433/44 = 1.6828

The F ratio is determined for each variable by dividing its MS by the MS error. Consult Chart 2 on pages 40 and 41 to see if the F ratio is significant. This procedure is described on page 22. In our example, the main effects are significantly different at the 1 percent level.

Use Tukey's Test to determine which samples are different. Because Duncan's Multiple Range Test (7) has previously been used and explained in this booklet (see pages 22-23), it is used here to determine which samples are different.

| | 0 | 150 | 200 | 250 |
|---|---|---|---|---|
| Average preferences | $\hat{\alpha}_0$ | $\hat{\alpha}_{150}$ | $\hat{\alpha}_{200}$ | $\hat{\alpha}_{250}$ |
| | 0.75 | -0.15625 | -0.3125 | -0.28125 |
| | A | B | C | D |
| | $\hat{\alpha}_0$ | $\hat{\alpha}_{150}$ | $\hat{\alpha}_{250}$ | $\hat{\alpha}_{200}$ |
| | 0.75 | -0.15625 | -0.28125 | -0.3125 |

S.E. $= \sqrt{}$ (MS error/Number of judgments for each sample) $= \sqrt{}$ (1.6828/24)
$= \sqrt{}$ 0.070 = 0.27

| P | 2 | 3 | 4 |
|---|---|---|---|
| rp (5 percent) | 2.86 | 3.01 | 3.10 |
| Rp | 0.77 | 0.81 | 0.84 |

$\hat{\alpha}_0 - \hat{\alpha}_{200} = 0.75 - (-0.3125) = 1.0625 > 0.84$

$\hat{\alpha}_0 - \hat{\alpha}_{250} = 0.75 - (-0.28125) = 1.03125 > 0.81$

$\hat{\alpha}_0 - \hat{\alpha}_{150} = 0.75 - (-0.15625) = 0.90625 > 0.77$

$\hat{\alpha}_{150} - \hat{\alpha}_{200} = -0.15625 - (-0.3125) = 0.15626 < 0.81$

$\hat{\alpha}_0$ $\hat{\alpha}_{150}$ $\hat{\alpha}_{250}$ $\hat{\alpha}_{200}$

The control sample is significantly different from the other samples. Its score is higher and therefore it would be considered to have better texture. This does not mean necessarily that the other samples are of poor texture. To determine the quality of the samples, the ratings given them by the panelists were tabulated to see if they were considered to be good, fair, or poor. An average score can be determined for each sample by assigning the values good = 3, fair = 2, poor = 1, and computing the average for each sample.

## HEDONIC SCALE

## SCORING

DATE _____ TASTER _____

PRODUCT _____

Taste test these samples and check how much you like or dislike each one. Use the appropriate scale to show your attitude by checking at the point that best describes your feeling about the sample. Please give a reason for this attitude. Remember you are the only one who can tell what you like. An honest expression of your personal feeling will help us.

| CODE 459 | CODE 667 | CODE 619 | CODE 347 |
|---|---|---|---|
| ___ Like extremely | ___ Like extremely | ___ Like extremely | ___ Like extremely |
| ___ Like very much | ___ Like very much | ___ Like very much | ___ Like very much |
| ___ Like moderately | ___ Like moderately | ___ Like moderately | ___ Like moderately |
| ___ Like slightly | ___ Like slightly | ___ Like slightly | ___ Like slightly |
| ___ Neither like nor dislike | ___ Neither like nor dislike | ___ Neither like nor dislike | ___ Neither like nor dislike |
| ___ Dislike slightly | ___ Dislike slightly | ___ Dislike slightly | ___ Dislike slightly |
| ___ Dislike moderately | ___ Dislike moderately | ___ Dislike moderately | ___ Dislike moderately |
| ___ Dislike very much | ___ Dislike very much | ___ Dislike very much | ___ Dislike very much |
| ___ Dislike extremely | ___ Dislike extremely | ___ Dislike extremely | ___ Dislike extremely |
| REASON | REASON | REASON | REASON |
| | | | |

Note: To analyze the results of this test use analysis of variance (see pages 20-23).

Numerical values: like extremely = 9
dislike extremely = 1

## PAIRED COMPARISON

## PREFERENCE

DATE _____ TASTER _____

PRODUCT _____

INSTRUCTIONS: (A) Here are two samples for evaluation. Please indicate which sample you prefer.

622 244

(B) Indicate the degree of preference between the two samples.

Slight _____

Moderate _____

Much _____

Extreme _____

Example:

To determine which of two samples of Cheddar cheese (Number 1 or Number 2) had more acceptable flavor, a paired comparison test was used. The samples were coded and one sample of each cheese was presented on each tray. Eight panelists were presented with three trays each, one after the other. This made a total of 24 judgments.

Cheese Number 2 was preferred 17 times in the 24 judgments. According to Statistical Chart 1 of Appendix II, on page 39, in the two-sample test, one sample must be preferred 18 times to be significantly more acceptable. So we conclude that neither cheese was significantly more acceptable for flavor.

## RANKING

## PREFERENCE

NAME _____

DATE _____

PRODUCT _____

Please rank these samples according to your preference.

| | Code |
|---|---|
| First | _____ |
| Second | _____ |
| Third | _____ |
| Fourth | _____ |

Note: The results are analyzed by converting ranks to scores and conducting analysis of variance as shown for Ranking Difference Analysis, page 24.

## APPENDIX II
## STATISTICAL CHART 1

| Number of tasters | Two-sample test, number of concurring choices necessary to establish significance | | | Triangle test difference analysis, number of correct answers necessary to establish significance | | |
|---|---|---|---|---|---|---|
| | * | ** | *** | * | ** | *** |
| 1 | – | – | – | – | – | – |
| 2 | – | – | – | – | – | – |
| 3 | – | – | – | 3 | – | – |
| 4 | – | – | – | 4 | – | – |
| 5 | – | – | – | 5 | 5 | – |
| 6 | 6 | – | – | 5 | 6 | – |
| 7 | 7 | – | – | 5 | 6 | 7 |
| 8 | 8 | 8 | – | 6 | 7 | 8 |
| 9 | 8 | 9 | – | 6 | 7 | 8 |
| 10 | 9 | 10 | – | 7 | 8 | 9 |
| 11 | 10 | 11 | 11 | 7 | 8 | 10 |
| 12 | 10 | 11 | 12 | 8 | 9 | 10 |
| 13 | 11 | 12 | 13 | 8 | 9 | 11 |
| 14 | 12 | 13 | 14 | 9 | 10 | 11 |
| 15 | 12 | 13 | 14 | 9 | 10 | 12 |
| 16 | 13 | 14 | 15 | 9 | 11 | 12 |
| 17 | 13 | 15 | 16 | 10 | 11 | 13 |
| 18 | 14 | 15 | 17 | 10 | 12 | 13 |
| 19 | 15 | 16 | 17 | 11 | 13 | 14 |
| 20 | 15 | 17 | 18 | 11 | 13 | 14 |
| 21 | 16 | 17 | 19 | 12 | 13 | 15 |
| 22 | 17 | 18 | 19 | 12 | 14 | 15 |
| 23 | 17 | 19 | 20 | 12 | 14 | 16 |
| 24 | 18 | 19 | 21 | 13 | 15 | 16 |
| 25 | 18 | 20 | 21 | 13 | 15 | 17 |
| 26 | 19 | 20 | 22 | 14 | 15 | 17 |
| 27 | 20 | 21 | 23 | 14 | 16 | 18 |
| 28 | 20 | 22 | 23 | 15 | 16 | 18 |
| 29 | 21 | 22 | 24 | 15 | 17 | 19 |
| 30 | 21 | 23 | 25 | 15 | 17 | 19 |
| 31 | 22 | 24 | 25 | 16 | 18 | 20 |
| 32 | 23 | 24 | 27 | 16 | 18 | 20 |
| 33 | 23 | 25 | 27 | 17 | 18 | 21 |
| 34 | 24 | 25 | 27 | 17 | 19 | 21 |
| 35 | 24 | 26 | 28 | 17 | 19 | 22 |
| 36 | 25 | 27 | 29 | 18 | 20 | 22 |
| 37 | 25 | 27 | 29 | 18 | 20 | 22 |
| 38 | 26 | 28 | 30 | 19 | 21 | 23 |
| 39 | 27 | 28 | 31 | 19 | 21 | 23 |
| 40 | 27 | 29 | 31 | 19 | 21 | 24 |
| 41 | 27 | 29 | 32 | 20 | 22 | 24 |
| 42 | 28 | 30 | 32 | 20 | 22 | 25 |
| 43 | 28 | 30 | 33 | 21 | 23 | 25 |
| 44 | 29 | 31 | 33 | 21 | 23 | 25 |
| 45 | 30 | 32 | 34 | 22 | 24 | 26 |
| 46 | 30 | 32 | 35 | 22 | 24 | 26 |
| 47 | 31 | 33 | 35 | 23 | 24 | 27 |
| 48 | 31 | 33 | 36 | 23 | 25 | 27 |
| 49 | 32 | 34 | 37 | 23 | 25 | 28 |
| 50 | 32 | 35 | 37 | 24 | 26 | 28 |

* 5 percent level of significance. ** 1 percent level. *** 0.1 percent level.

## STATISTICAL CHART 2

Variance Ratio – 5 Percent Points for Distribution of F

$n_1$ – Degrees of freedom for numerator

$n_2$ – Degrees of freedom for denominator

| $n_2$ \ $n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 12 | 24 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 161.4 | 199.5 | 215.7 | 224.6 | 230.2 | 234.0 | 238.9 | 243.9 | 249.0 | 254.3 |
| 2 | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.37 | 19.41 | 19.45 | 19.50 |
| 3 | 10.13 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.84 | 8.74 | 8.64 | 8.53 |
| 4 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.04 | 5.91 | 5.77 | 5.63 |
| 5 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.82 | 4.68 | 4.53 | 4.36 |
| 6 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.15 | 4.00 | 3.84 | 3.67 |
| 7 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.73 | 3.57 | 3.41 | 3.23 |
| 8 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.44 | 3.28 | 3.12 | 2.93 |
| 9 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.23 | 3.07 | 2.90 | 2.71 |
| 10 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.07 | 2.91 | 2.74 | 2.54 |
| 11 | 4.84 | 3.98 | 3.59 | 3.36 | 3.20 | 3.09 | 2.95 | 2.79 | 2.61 | 2.40 |
| 12 | 4.75 | 3.88 | 3.49 | 3.26 | 3.11 | 3.00 | 2.85 | 2.69 | 2.50 | 2.30 |
| 13 | 4.67 | 3.80 | 3.41 | 3.18 | 3.02 | 2.92 | 2.77 | 2.60 | 2.42 | 2.21 |
| 14 | 4.60 | 3.74 | 3.34 | 3.11 | 2.96 | 2.85 | 2.70 | 2.53 | 2.35 | 2.13 |
| 15 | 4.54 | 3.68 | 3.29 | 3.06 | 2.90 | 2.79 | 2.64 | 2.48 | 2.29 | 2.07 |
| 16 | 4.49 | 3.63 | 3.24 | 3.01 | 2.85 | 2.74 | 2.59 | 2.42 | 2.24 | 2.01 |
| 17 | 4.45 | 3.59 | 3.20 | 2.96 | 2.81 | 2.70 | 2.55 | 2.38 | 2.19 | 1.96 |
| 18 | 4.41 | 3.55 | 3.16 | 2.93 | 2.77 | 2.66 | 2.51 | 2.34 | 2.15 | 1.92 |
| 19 | 4.38 | 3.52 | 3.13 | 2.90 | 2.74 | 2.63 | 2.48 | 2.31 | 2.11 | 1.88 |
| 20 | 4.35 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.45 | 2.28 | 2.08 | 1.84 |
| 21 | 4.32 | 3.47 | 3.07 | 2.84 | 2.68 | 2.57 | 2.42 | 2.25 | 2.05 | 1.81 |
| 22 | 4.30 | 3.44 | 3.05 | 2.82 | 2.66 | 2.55 | 2.40 | 2.23 | 2.03 | 1.78 |
| 23 | 4.28 | 3.42 | 3.03 | 2.80 | 2.64 | 2.53 | 2.38 | 2.20 | 2.00 | 1.76 |
| 24 | 4.26 | 3.40 | 3.01 | 2.78 | 2.62 | 2.51 | 2.36 | 2.18 | 1.98 | 1.73 |
| 25 | 4.24 | 3.38 | 2.99 | 2.76 | 2.60 | 2.49 | 2.34 | 2.16 | 1.96 | 1.71 |
| 26 | 4.22 | 3.37 | 2.98 | 2.74 | 2.59 | 2.47 | 2.32 | 2.15 | 1.95 | 1.69 |
| 27 | 4.21 | 3.35 | 2.96 | 2.73 | 2.57 | 2.46 | 2.30 | 2.13 | 1.93 | 1.67 |
| 28 | 4.20 | 3.34 | 2.95 | 2.71 | 2.56 | 2.44 | 2.29 | 2.12 | 1.91 | 1.65 |
| 29 | 4.18 | 3.33 | 2.93 | 2.70 | 2.54 | 2.43 | 2.28 | 2.10 | 1.90 | 1.64 |
| 30 | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.27 | 2.09 | 1.89 | 1.62 |
| 40 | 4.08 | 3.23 | 2.84 | 2.61 | 2.45 | 2.34 | 2.18 | 2.00 | 1.79 | 1.51 |
| 60 | 4.00 | 3.15 | 2.76 | 2.52 | 2.37 | 2.25 | 2.10 | 1.92 | 1.70 | 1.39 |
| 120 | 3.92 | 3.07 | 2.68 | 2.45 | 2.29 | 2.17 | 2.02 | 1.83 | 1.61 | 1.25 |
| ∞ | 3.84 | 2.99 | 2.60 | 2.37 | 2.21 | 2.09 | 1.94 | 1.75 | 1.52 | 1.00 |

## STATISTICAL CHART 2 – Concluded

Variance Ratio – 1 Percent Points for Distribution of F

$n_1$ – Degrees of freedom for numerator

$n_2$ – Degrees of freedom for denominator

| $n_2$ \ $n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 12 | 24 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4052 | 4999 | 5403 | 5625 | 5764 | 5859 | 5981 | 6106 | 6234 | 6366 |
| 2 | 98.49 | 99.00 | 99.17 | 99.25 | 99.30 | 99.33 | 99.36 | 99.42 | 99.46 | 99.50 |
| 3 | 34.12 | 30.81 | 29.46 | 28.71 | 28.24 | 27.91 | 27.49 | 27.05 | 26.60 | 26.12 |
| 4 | 21.20 | 18.00 | 16.69 | 15.98 | 15.52 | 15.21 | 14.80 | 14.37 | 13.93 | 13.46 |
| 5 | 16.26 | 13.27 | 12.06 | 11.39 | 10.97 | 10.67 | 10.29 | 9.89 | 9.47 | 9.02 |
| 6 | 13.74 | 10.92 | 9.78 | 9.15 | 8.75 | 8.47 | 8.10 | 7.72 | 7.31 | 6.88 |
| 7 | 12.25 | 9.55 | 8.45 | 7.85 | 7.46 | 7.19 | 6.84 | 6.47 | 6.07 | 5.65 |
| 8 | 11.26 | 8.65 | 7.59 | 7.01 | 6.63 | 6.37 | 6.03 | 5.67 | 5.28 | 4.86 |
| 9 | 10.56 | 8.02 | 6.99 | 6.42 | 6.06 | 5.80 | 5.47 | 5.11 | 4.73 | 4.31 |
| 10 | 10.04 | 7.56 | 6.55 | 5.99 | 5.64 | 5.39 | 5.06 | 4.71 | 4.33 | 3.91 |
| 11 | 9.65 | 7.20 | 6.22 | 5.67 | 5.32 | 5.07 | 4.74 | 4.40 | 4.02 | 3.60 |
| 12 | 9.33 | 6.93 | 5.95 | 5.41 | 5.06 | 4.82 | 4.50 | 4.16 | 3.78 | 3.36 |
| 13 | 9.07 | 6.70 | 5.74 | 5.20 | 4.86 | 4.62 | 4.30 | 3.96 | 3.59 | 3.16 |
| 14 | 8.86 | 6.51 | 5.56 | 5.03 | 4.69 | 4.46 | 4.14 | 3.80 | 3.43 | 3.00 |
| 15 | 8.68 | 6.36 | 5.42 | 4.89 | 4.56 | 4.32 | 4.00 | 3.67 | 3.29 | 2.87 |
| 16 | 8.53 | 6.23 | 5.29 | 4.77 | 4.44 | 4.20 | 3.89 | 3.55 | 3.18 | 2.75 |
| 17 | 8.40 | 6.11 | 5.18 | 4.67 | 4.34 | 4.10 | 3.79 | 3.45 | 3.08 | 2.65 |
| 18 | 8.28 | 6.01 | 5.09 | 4.58 | 4.25 | 4.01 | 3.71 | 3.37 | 3.00 | 2.57 |
| 19 | 8.18 | 5.93 | 5.01 | 4.50 | 4.17 | 3.94 | 3.63 | 3.30 | 2.92 | 2.49 |
| 20 | 8.10 | 5.85 | 4.94 | 4.43 | 4.10 | 3.87 | 3.56 | 3.23 | 2.86 | 2.42 |
| 21 | 8.02 | 5.78 | 4.87 | 4.37 | 4.04 | 3.81 | 3.51 | 3.17 | 2.80 | 2.36 |
| 22 | 7.94 | 5.72 | 4.82 | 4.31 | 3.99 | 3.76 | 3.45 | 3.12 | 2.75 | 2.31 |
| 23 | 7.88 | 5.66 | 4.76 | 4.26 | 3.94 | 3.71 | 3.41 | 3.07 | 2.70 | 2.26 |
| 24 | 7.82 | 5.61 | 4.72 | 4.22 | 3.90 | 3.67 | 3.36 | 3.03 | 2.66 | 2.21 |
| 25 | 7.77 | 5.57 | 4.68 | 4.18 | 3.86 | 3.63 | 3.32 | 2.99 | 2.62 | 2.17 |
| 26 | 7.72 | 5.53 | 4.64 | 4.14 | 3.82 | 3.59 | 3.29 | 2.96 | 2.58 | 2.13 |
| 27 | 7.68 | 5.49 | 4.60 | 4.11 | 3.78 | 3.56 | 3.26 | 2.93 | 2.55 | 2.10 |
| 28 | 7.64 | 5.45 | 4.57 | 4.07 | 3.75 | 3.53 | 3.23 | 2.90 | 2.52 | 2.06 |
| 29 | 7.60 | 5.42 | 4.54 | 4.04 | 3.73 | 3.50 | 3.20 | 2.87 | 2.49 | 2.03 |
| 30 | 7.56 | 5.39 | 4.51 | 4.02 | 3.70 | 3.47 | 3.17 | 2.84 | 2.47 | 2.01 |
| 40 | 7.31 | 5.18 | 4.31 | 3.83 | 3.51 | 3.29 | 2.99 | 2.66 | 2.29 | 1.80 |
| 60 | 7.08 | 4.98 | 4.13 | 3.65 | 3.34 | 3.12 | 2.82 | 2.50 | 2.12 | 1.60 |
| 120 | 6.85 | 4.79 | 3.95 | 3.48 | 3.17 | 2.96 | 2.66 | 2.34 | 1.95 | 1.38 |
| ∞ | 6.64 | 4.60 | 3.78 | 3.32 | 3.02 | 2.80 | 2.51 | 2.18 | 1.79 | 1.00 |

# STATISTICAL CHART 3

Multiple F Tests
Significant Studentized Ranges for a 5 Percent Level –
Multiple Range Test

| n \ p | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 14 | 16 | 18 | 20 | 50 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 | 18.0 |
| 2 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 | 6.09 |
| 3 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 | 4.50 |
| 4 | 3.93 | 4.01 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 | 4.02 |
| 5 | 3.64 | 3.74 | 3.79 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 | 3.83 |
| 6 | 3.46 | 3.58 | 3.64 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 | 3.68 |
| 7 | 3.35 | 3.47 | 3.54 | 3.58 | 3.60 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 | 3.61 |
| 8 | 3.26 | 3.39 | 3.47 | 3.52 | 3.55 | 3.56 | 3.56 | 3.56 | 3.56 | 3.56 | 3.56 | 3.56 | 3.56 | 3.56 | 3.56 | 3.56 |
| 9 | 3.20 | 3.34 | 3.41 | 3.47 | 3.50 | 3.52 | 3.52 | 3.52 | 3.52 | 3.52 | 3.52 | 3.52 | 3.52 | 3.52 | 3.52 | 3.52 |
| 10 | 3.15 | 3.30 | 3.37 | 3.43 | 3.46 | 3.47 | 3.47 | 3.47 | 3.47 | 3.47 | 3.47 | 3.47 | 3.47 | 3.48 | 3.48 | 3.48 |
| 11 | 3.11 | 3.27 | 3.35 | 3.39 | 3.43 | 3.44 | 3.45 | 3.46 | 3.46 | 3.46 | 3.46 | 3.46 | 3.47 | 3.48 | 3.48 | 3.48 |
| 12 | 3.08 | 3.23 | 3.33 | 3.36 | 3.40 | 3.42 | 3.44 | 3.44 | 3.46 | 3.46 | 3.46 | 3.46 | 3.47 | 3.48 | 3.48 | 3.48 |
| 13 | 3.06 | 3.21 | 3.30 | 3.35 | 3.38 | 3.41 | 3.42 | 3.44 | 3.45 | 3.45 | 3.46 | 3.46 | 3.47 | 3.47 | 3.47 | 3.47 |
| 14 | 3.03 | 3.18 | 3.27 | 3.33 | 3.37 | 3.39 | 3.41 | 3.42 | 3.44 | 3.45 | 3.46 | 3.46 | 3.47 | 3.47 | 3.47 | 3.47 |
| 15 | 3.01 | 3.16 | 3.25 | 3.31 | 3.36 | 3.38 | 3.40 | 3.42 | 3.43 | 3.44 | 3.45 | 3.46 | 3.47 | 3.47 | 3.47 | 3.47 |
| 16 | 3.00 | 3.15 | 3.23 | 3.30 | 3.34 | 3.37 | 3.30 | 3.41 | 3.43 | 3.44 | 3.45 | 3.46 | 3.47 | 3.47 | 3.47 | 3.47 |
| 17 | 2.98 | 3.13 | 3.22 | 3.28 | 3.33 | 3.36 | 3.38 | 3.40 | 3.42 | 3.44 | 3.45 | 3.46 | 3.47 | 3.47 | 3.47 | 3.47 |
| 18 | 2.97 | 3.12 | 3.21 | 3.27 | 3.32 | 3.35 | 3.37 | 3.39 | 3.41 | 3.43 | 3.45 | 3.46 | 3.47 | 3.47 | 3.47 | 3.47 |
| 19 | 2.96 | 3.11 | 3.19 | 3.26 | 3.31 | 3.35 | 3.37 | 3.39 | 3.41 | 3.43 | 3.44 | 3.46 | 3.47 | 3.47 | 3.47 | 3.47 |
| 20 | 2.95 | 3.10 | 3.18 | 3.25 | 3.30 | 3.34 | 3.36 | 3.38 | 3.40 | 3.43 | 3.44 | 3.46 | 3.46 | 3.47 | 3.47 | 3.47 |
| 22 | 2.93 | 3.08 | 3.17 | 3.24 | 3.29 | 3.32 | 3.35 | 3.37 | 3.39 | 3.42 | 3.44 | 3.45 | 3.46 | 3.47 | 3.47 | 3.47 |
| 24 | 2.92 | 3.07 | 3.15 | 3.22 | 3.28 | 3.31 | 3.34 | 3.37 | 3.38 | 3.41 | 3.44 | 3.45 | 3.46 | 3.47 | 3.47 | 3.47 |
| 26 | 2.91 | 3.06 | 3.14 | 3.21 | 3.27 | 3.30 | 3.34 | 3.36 | 3.38 | 3.41 | 3.43 | 3.45 | 3.46 | 3.47 | 3.47 | 3.47 |
| 28 | 2.90 | 3.04 | 3.13 | 3.20 | 3.26 | 3.30 | 3.33 | 3.35 | 3.37 | 3.40 | 3.43 | 3.45 | 3.46 | 3.47 | 3.47 | 3.47 |
| 30 | 2.89 | 3.04 | 3.12 | 3.20 | 3.25 | 3.29 | 3.32 | 3.35 | 3.37 | 3.40 | 3.43 | 3.44 | 3.46 | 3.47 | 3.47 | 3.47 |
| 40 | 2.86 | 3.01 | 3.10 | 3.17 | 3.22 | 3.27 | 3.30 | 3.33 | 3.35 | 3.39 | 3.42 | 3.44 | 3.46 | 3.47 | 3.47 | 3.47 |
| 60 | 2.83 | 2.98 | 3.08 | 3.14 | 3.20 | 3.24 | 3.28 | 3.31 | 3.33 | 3.37 | 3.40 | 3.43 | 3.45 | 3.47 | 3.48 | 3.48 |
| 100 | 2.80 | 2.95 | 3.05 | 3.12 | 3.18 | 3.22 | 3.26 | 3.29 | 3.32 | 3.36 | 3.40 | 3.42 | 3.45 | 3.47 | 3.53 | 3.53 |
| ∞ | 2.77 | 2.92 | 3.02 | 3.09 | 3.15 | 3.19 | 3.23 | 3.26 | 3.29 | 3.34 | 3.38 | 3.41 | 3.44 | 3.47 | 3.61 | 3.67 |

# STATISTICAL CHART 4

Multiple F Tests
Significant Studentized Ranges for a 1 Percent Level –
Multiple Range Test

| n \ p | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 14 | 16 | 18 | 20 | 50 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 | 90.0 |
| 2 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 | 14.0 |
| 3 | 8.26 | 8.5 | 8.6 | 8.7 | 8.8 | 8.9 | 8.9 | 9.0 | 9.0 | 9.0 | 9.1 | 9.2 | 9.3 | 9.3 | 9.3 | 9.3 |
| 4 | 6.31 | 6.8 | 6.9 | 7.0 | 7.1 | 7.1 | 7.2 | 7.2 | 7.3 | 7.3 | 7.4 | 7.4 | 7.5 | 7.5 | 7.5 | 7.5 |
| 5 | 5.70 | 5.96 | 6.11 | 6.18 | 6.26 | 6.33 | 6.40 | 6.44 | 6.5 | 6.6 | 6.6 | 6.7 | 6.7 | 6.8 | 6.8 | 6.8 |
| 6 | 5.24 | 5.51 | 5.65 | 5.73 | 5.81 | 5.88 | 5.95 | 6.00 | 6.0 | 6.1 | 6.2 | 6.2 | 6.3 | 6.3 | 6.3 | 6.3 |
| 7 | 4.93 | 5.22 | 5.37 | 5.45 | 5.53 | 5.61 | 5.69 | 5.73 | 5.8 | 5.8 | 5.9 | 5.9 | 6.0 | 6.0 | 6.0 | 6.0 |
| 8 | 4.74 | 5.00 | 5.14 | 5.23 | 5.32 | 5.40 | 5.47 | 5.51 | 5.5 | 5.6 | 5.7 | 5.7 | 5.8 | 5.8 | 5.8 | 5.8 |
| 9 | 4.60 | 4.86 | 4.99 | 5.08 | 5.17 | 5.25 | 5.32 | 5.36 | 5.4 | 5.5 | 5.5 | 5.6 | 5.7 | 5.7 | 5.7 | 5.7 |
| 10 | 4.48 | 4.73 | 4.88 | 4.96 | 5.06 | 5.13 | 5.20 | 5.24 | 5.28 | 5.36 | 5.42 | 5.48 | 5.54 | 5.55 | 5.55 | 5.55 |
| 11 | 4.39 | 4.63 | 4.77 | 4.86 | 4.94 | 5.01 | 5.06 | 5.12 | 5.15 | 5.24 | 5.28 | 5.34 | 5.38 | 5.39 | 5.39 | 5.39 |
| 12 | 4.32 | 4.55 | 4.68 | 4.76 | 4.84 | 4.92 | 4.96 | 5.02 | 5.07 | 5.13 | 5.17 | 5.22 | 5.24 | 5.26 | 5.26 | 5.26 |
| 13 | 4.26 | 4.48 | 4.62 | 4.69 | 4.74 | 4.84 | 4.88 | 4.94 | 4.98 | 5.04 | 5.08 | 5.13 | 5.14 | 5.15 | 5.15 | 5.15 |
| 14 | 4.21 | 4.42 | 4.55 | 4.63 | 4.70 | 4.78 | 4.83 | 4.87 | 4.91 | 4.96 | 5.00 | 5.04 | 5.06 | 5.07 | 5.07 | 5.07 |
| 15 | 4.17 | 4.37 | 4.50 | 4.58 | 4.64 | 4.72 | 4.77 | 4.81 | 4.84 | 4.90 | 4.94 | 4.97 | 4.99 | 5.00 | 5.00 | 5.00 |
| 16 | 4.13 | 4.34 | 4.45 | 4.54 | 4.60 | 4.67 | 4.72 | 4.76 | 4.79 | 4.84 | 4.88 | 4.91 | 4.93 | 4.94 | 4.94 | 4.94 |
| 17 | 4.10 | 4.30 | 4.41 | 4.50 | 4.56 | 4.63 | 4.68 | 4.72 | 4.75 | 4.80 | 4.83 | 4.86 | 4.88 | 4.89 | 4.89 | 4.89 |
| 18 | 4.07 | 4.27 | 4.38 | 4.46 | 4.53 | 4.59 | 4.64 | 4.68 | 4.71 | 4.76 | 4.79 | 4.82 | 4.84 | 4.85 | 4.85 | 4.85 |
| 19 | 4.05 | 4.24 | 4.35 | 4.43 | 4.50 | 4.56 | 4.61 | 4.64 | 4.67 | 4.72 | 4.76 | 4.79 | 4.81 | 4.82 | 4.82 | 4.82 |
| 20 | 4.02 | 4.22 | 4.33 | 4.40 | 4.47 | 4.53 | 4.58 | 4.61 | 4.65 | 4.69 | 4.73 | 4.76 | 4.78 | 4.79 | 4.79 | 4.79 |
| 22 | 3.99 | 4.17 | 4.28 | 4.36 | 4.42 | 4.48 | 4.53 | 4.57 | 4.60 | 4.65 | 4.68 | 4.71 | 4.74 | 4.75 | 4.75 | 4.75 |
| 24 | 3.96 | 4.14 | 4.24 | 4.33 | 4.39 | 4.44 | 4.49 | 4.53 | 4.57 | 4.62 | 4.64 | 4.67 | 4.70 | 4.72 | 4.74 | 4.74 |
| 26 | 3.93 | 4.11 | 4.21 | 4.30 | 4.36 | 4.41 | 4.46 | 4.50 | 4.53 | 4.58 | 4.62 | 4.65 | 4.67 | 4.69 | 4.73 | 4.73 |
| 28 | 3.91 | 4.08 | 4.18 | 4.28 | 4.34 | 4.39 | 4.43 | 4.47 | 4.51 | 4.56 | 4.60 | 4.62 | 4.65 | 4.67 | 4.72 | 4.72 |
| 30 | 3.89 | 4.06 | 4.16 | 4.22 | 4.32 | 4.36 | 4.41 | 4.45 | 4.48 | 4.54 | 4.58 | 4.61 | 4.63 | 4.65 | 4.71 | 4.71 |
| 40 | 3.82 | 3.99 | 4.10 | 4.17 | 4.24 | 4.30 | 4.34 | 4.37 | 4.41 | 4.46 | 4.51 | 4.54 | 4.57 | 4.59 | 4.69 | 4.69 |
| 60 | 3.76 | 3.92 | 4.03 | 4.12 | 4.17 | 4.23 | 4.27 | 4.31 | 4.34 | 4.39 | 4.44 | 4.47 | 4.50 | 4.53 | 4.66 | 4.66 |
| 100 | 3.71 | 3.86 | 3.98 | 4.06 | 4.11 | 4.17 | 4.21 | 4.25 | 4.29 | 4.35 | 4.38 | 4.42 | 4.45 | 4.48 | 4.64 | 4.65 |
| ∞ | 3.64 | 3.80 | 3.90 | 3.98 | 4.04 | 4.09 | 4.14 | 4.17 | 4.20 | 4.26 | 4.31 | 4.34 | 4.38 | 4.41 | 4.60 | 4.68 |

## STATISTICAL CHART 5

Scores for Ranked Data

The mean deviations of the 1st, 2nd, 3rd . . . largest members of samples of different sizes; zero and negative values omitted.

| Ordinal number | Size of Sample | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | | .56 | .85 | 1.03 | 1.16 | 1.27 | 1.35 | 1.42 | 1.49 | 1.54 |
| 2 | | | | .30 | .50 | .64 | .76 | .85 | .93 | 1.00 |
| 3 | | | | | | .20 | .35 | .47 | .57 | .66 |
| 4 | | | | | | | | .15 | .27 | .38 |
| 5 | | | | | | | | | | .12 |
| | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 1 | 1.59 | 1.63 | 1.67 | 1.70 | 1.74 | 1.76 | 1.79 | 1.82 | 1.84 | 1.87 |
| 2 | 1.06 | 1.12 | 1.16 | 1.21 | 1.25 | 1.28 | 1.32 | 1.35 | 1.38 | 1.41 |
| 3 | .73 | .79 | .85 | .90 | .95 | .99 | 1.03 | 1.07 | 1.10 | 1.13 |
| 4 | .46 | .54 | .60 | .66 | .71 | .76 | .81 | .85 | .89 | .92 |
| 5 | .22 | .31 | .39 | .46 | .52 | .57 | .62 | .67 | .71 | .75 |
| 6 | | .10 | .19 | .27 | .34 | .39 | .45 | .50 | .55 | .59 |
| 7 | | | | .09 | .17 | .23 | .30 | .35 | .40 | .45 |
| 8 | | | | | | .08 | .15 | .21 | .26 | .31 |
| 9 | | | | | | | | .07 | .13 | .19 |
| 10 | | | | | | | | | | .06 |
| | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| 1 | 1.89 | 1.91 | 1.93 | 1.95 | 1.97 | 1.98 | 2.00 | 2.01 | 2.03 | 2.04 |
| 2 | 1.43 | 1.46 | 1.48 | 1.50 | 1.52 | 1.54 | 1.56 | 1.58 | 1.60 | 1.62 |
| 3 | 1.16 | 1.19 | 1.21 | 1.24 | 1.26 | 1.29 | 1.31 | 1.33 | 1.35 | 1.36 |
| 4 | .95 | .98 | 1.01 | 1.04 | 1.07 | 1.09 | 1.11 | 1.14 | 1.16 | 1.18 |
| 5 | .78 | .82 | .85 | .88 | .91 | .93 | .96 | .98 | 1.00 | 1.03 |
| 6 | .63 | .67 | .70 | .73 | .76 | .79 | .82 | .85 | .87 | .89 |
| 7 | .49 | .53 | .57 | .60 | .64 | .67 | .70 | .73 | .75 | .78 |
| 8 | .36 | .41 | .45 | .48 | .52 | .55 | .58 | .61 | .64 | .67 |
| 9 | .24 | .29 | .33 | .37 | .41 | .44 | .48 | .51 | .54 | .57 |
| 10 | .12 | .17 | .22 | .26 | .30 | .34 | .38 | .41 | .44 | .47 |
| 11 | | .06 | .11 | .16 | .20 | .24 | .28 | .32 | .35 | .38 |
| 12 | | | | .05 | .10 | .14 | .19 | .22 | .26 | .29 |
| 13 | | | | | | .05 | .09 | .13 | .17 | .21 |
| 14 | | | | | | | | .04 | .09 | .12 |
| 15 | | | | | | | | | | .04 |

**Tests of psychological preference and some other experimental data suffice to place a series of magnitudes in order of preference, without supplying metrical values. Analyses of variance, correlations, etc., can be carried out on such data by using the normal scores, appropriate to each position in order, in a sample of the size observed. Ties may be scored with the means of the ordinal values involved, but in such cases the sums of squares will require correction.**

# APPENDIX III

## NOTES ON INTRODUCTORY STATISTICS[1]

The purpose of these notes is to put in a written form the content of a 3-hour discussion on some basic statistical concepts, held at the Food Research Institute of the Department of Agriculture.

The term statistics includes the numerical descriptions (figures) of the quantitative aspects of things, and the body of methods (subject) used for making decisions in the face of uncertainty. The decisions may vary. They may be concerned with estimating the probability of rain on a summer day after considering the available meteorological data, whether or not to accept a shipment of parts after only partial inspection, or how to set up an experimental plan to test several varieties of geese for tenderness. These notes deal with the concept of statistics as a body of methods.

Statistics is helpful to research workers in many ways, such as by suggesting which to observe and how to observe it (Theory of Design of Experiments and Theory of Sampling) and how to summarize results in forms that are comprehensible (Descriptive Statistics). Owing to experimental error (chance circumstance), data and predictions cannot be expected to agree exactly even if the scientist's theories are correct. Therefore, when mathematical results from the Theory of Chance are applied to statistical problems (Inferential Statistics), they help to draw conclusions from the data.

### SAMPLE AND POPULATION

Sample (data, observations) is the "number" that have been observed. Population is the totality of all possible observations of the same kind.

Sample results vary by chance, and the pattern of chance variation depends on the population from which the samples have been drawn. A sample is not a miniature replicate of the population, so when decisions about a population are based on a sample, it is necessary to make allowance for the role of chance.

### RANDOM SAMPLE

If a sample is to be representative of the population, *each member* of the population *should have an equal chance* of being included in the sample. But this requirement is not enough in itself to make up a random sample.

---

[1] This appendix was prepared by Andres Petrasovits, Statistical Research Service, Canada Department of Agriculture, Ottawa.

A sample of size N is said to be random if *each combination* of N items in the population has an equal chance of being chosen.

To stress the importance of the italicized words in the above definition, consider the following:

Example: A college catalogue of students has 50 pages. A sample of 50 students was taken by selecting at random one student from each page. Note that this is not a random sample, because a sample including two students listed on the same page has a zero chance of being chosen.

It must be pointed out that samples that are not strictly random are often preferable to random samples on the grounds of convenience or of increased precision (Stratification).

## PARAMETER AND STATISTIC

Samples are taken to learn about the population being sampled. A parameter is a quantitative characteristic of the population. A statistic is a mathematical function of the sample values or a value computed from the sample.

Example 1

Suppose we are interested in learning about the average weight of adult Canadians. The true average weight (call it $\mu$) of Canadians is clearly impossible (physically and economically) to compute. It is possible, however, to take a sample of 100 Canadians, register their weights, and compute the mean of the sample weights. Let us denote this mean obtained from the sample $\overline{X}$, and suppose that, in this case, we obtained $\overline{X} = 141.5$ pounds. As a first approximation it would be reasonable to regard 141.5 pounds as an estimate of $\mu$, the true but unknown average weight for all Canadians.

Then we say: $\overline{X}$ is a statistic computed from the sample, $\mu$ is a parameter of the population, and $\overline{X}$ is an estimate of $\mu$.

Example 2

If we are interested in the variability of weights, an indicator of dispersion that is often used is the range (largest minus smallest). As with the mean weights, the true range (call it R) of adult Canadian body weights cannot be practically obtained, but we can use the sample of size 100 (from example 1) to obtain an estimate of the true range on the basis of the range observed in the sample. Let us denote the latter by r.

Then we say: r is a statistic computed from the sample and R is a parameter of the population: r is an estimate of R.

Example 3

A political poll is to be taken for Ontario to estimate the proportion of Liberals. Clearly, the true proportion of Liberals (call it $\pi$) would be practically impossible to compute. We can, however, obtain the proportion of Liberals from the sample taken (call it p) and regard p as an estimate of $\pi$. Suppose that p was found to be 0.46, then we say that p = 0.46 is a statistic computed from the sample, $\pi$ is a parameter of the population, and p is an estimate of $\pi$.

It should be pointed out that many statistics can be computed from a sample. In example 1, other statistics besides $\overline{X}$ may give an estimate of $\mu$, for instance the most frequent value (mode) or the average of the extremes.

## POPULATION OF SAMPLE MEANS

Consider the following finite population P: (1, 2, 3, 4). Note that two summary measures of interest about P are the population mean $\mu = 2.5$ and the population range $R = 4 - 1 = 3$.

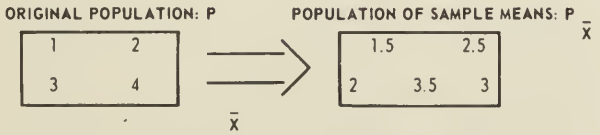
Let us form all possible samples of size 2, say, without replacement from P and for each sample let us compute the sample mean. We have:

| Sample | $\overline{X}$ |
|---|---|
| (1, 2) | 1.5 |
| (1, 3) | 2 |
| (1, 4) | 2.5 |
| (2, 3) | 2.5 |
| (2, 4) | 3 |
| (3, 4) | 3.5 |

The set of sample means

$P_{\overline{X}} = (1.5, 2, 2.5, 2.5, 3, 3.5)$

is called the population of sample means of size 2 obtained from the original population P.




When comparing the original population, P, with the population of sample means of size 2, $P_{\overline{X}}$, it should be noticed that:

(i) the mean of the population of sample means is equal to the mean of the original population,

(ii) the variability (dispersion) in the population of sample means is smaller than the variability in the original population.

The above results are illustrated by the previous example.

For (i),

$$(1 + 2 + 3 + 4)/4 = (1.5 + 2 + 2.5 + 2.5 + 3 + 3.5)/6 = 2.5,$$

and for (ii), if we agree to use the range as a measure of variability, then

range for original population = 4 − 1 = 3, and
range for population of sample means = 3.5 − 1.5 = 2.

As an exercise for the reader it may be useful to examine the population of sample means derived by taking all possible samples of size 3 from the population:

$$P = (2, 3, 3, 7, 7, 0).$$

## NORMAL DISTRIBUTION

### Frequency Distribution

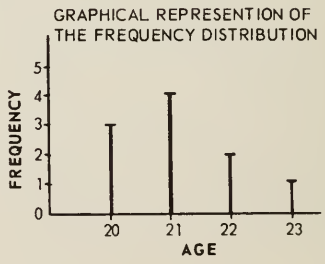
The pattern (characteristics) of a population or sample is shown by grouping the data in a frequency table. This table is called the frequency distribution of the population or the sample.

Example: The ages in years of 10 students in a classroom were:

21, 20, 21, 22, 20, 22, 21, 20, 23, 21.

Frequency distribution

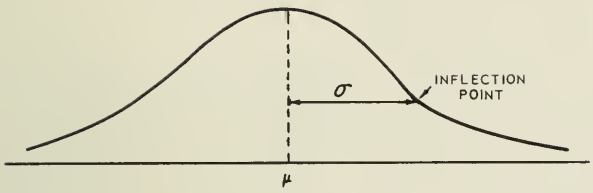
| Age | Frequency |
|---|---|
| 20 | 3 |
| 21 | 4 |
| 22 | 2 |
| 23 | 1 |




### Characterization of the Normal Distribution

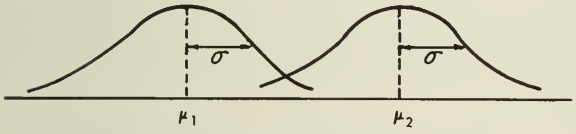
Roughly speaking, a normal distribution is a frequency distribution whose graphical description is "bell-shaped." It must be noted that not all bell-shaped distributions are normal.

A normal distribution is fully determined by two parameters: the mean ($\mu$) and the standard deviation ($\sigma$). A full definition of the latter is available in any introductory statistics textbook.

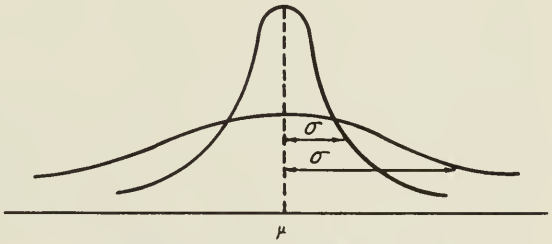



The physical meaning of the quantities $\mu$ and $\sigma$.

The mean, $\mu$, establishes the "location" or "center" of the distribution. The standard deviation, $\sigma$, measures dispersion and is given by the horizontal distance from $\mu$ to the inflection points of the normal curve. An important characteristic of the normal distribution is its symmetry about the mean $\mu$.




NORMAL CURVES WITH DIFFERENT MEANS (LOCATIONS) AND SAME STANDARD DEVIATION (DISPERSION)




NORMAL CURVES WITH SAME MEAN (LOCATION) AND DIFFERENT STANDARD DEVIATIONS

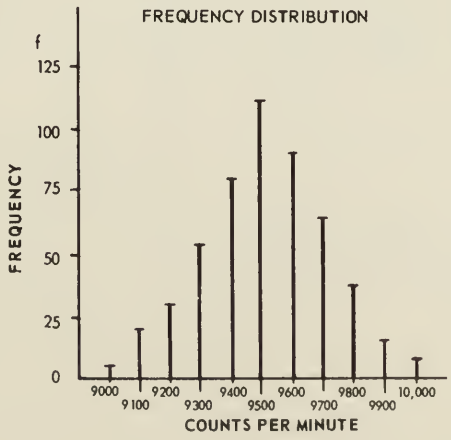
These figures show how the mean ($\mu$) and the standard deviation ($\sigma$) affect the appearance of the normal distribution.

Example

In a study of radioactivity, measurements and counts per minute were registered with their frequency.

Frequency distribution

| Counts per minute | Frequency |
|---|---|
| 9000 | 3 |
| 9100 | 20 |
| 9200 | 31 |
| 9300 | 54 |
| 9400 | 81 |
| 9500 | 112 |
| 9600 | 88 |
| 9700 | 64 |
| 9800 | 37 |
| 9900 | 14 |
| 10000 | 6 |



## CENTRAL LIMIT THEOREM

*The normal distribution arises* from some natural populations (for example, radioactivity counts) and *from all populations of sample means* (of large enough sample size). This is an important fact in the Theory of Statistics and it is known as the Central Limit Theorem, which states that as the sample size increases, the distribution of the population of sample means becomes more like the "bell-shaped" normal distribution regardless of the shape of the distribution of the original population.

Example: Consider the population:

$P_1 = (0, 1, 2, 3, 4, 5)$

and construct from it the population of sample means of size 2 ($P_2$) and the population of sample means of size 4 ($P_4$). From the frequency distributions of $P_1$, $P_2$, and $P_4$ construct graphical representations. It will be observed that the shape of the graphical representation of $P_2$ looks more bell-shaped than that of $P_1$ and that the shape of the graphical representation of $P_4$ looks more bell-shaped than that of $P_2$.

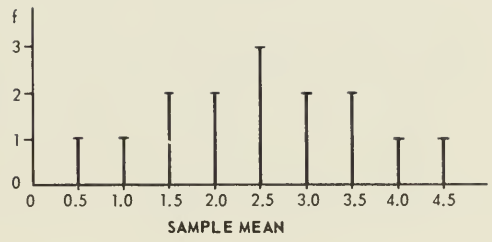
| Construction of $P_2$ | | | | Construction of $P_4$ | | | |
|---|---|---|---|---|---|---|---|
| Sample | $\bar{X}$ | Sample | $\bar{X}$ | Sample | $\bar{X}$ | Sample | $\bar{X}$ |
| (0, 1) | 0.5 | (1, 4) | 2.5 | (0, 1, 2, 3) | 1.50 | (0, 2, 3, 5) | 2.50 |
| (0, 2) | 1.0 | (0, 5) | 3.0 | (0, 1, 2, 4) | 1.75 | (0, 2, 4, 5) | 2.75 |
| (0, 3) | 1.5 | (2, 3) | 2.5 | (0, 1, 2, 5) | 2.00 | (0, 3, 4, 5) | 3.00 |
| (0, 4) | 2.0 | (2, 4) | 3.0 | (0, 1, 3, 4) | 2.00 | (1, 2, 3, 4) | 2.50 |
| (0, 5) | 2.5 | (2, 5) | 3.5 | (0, 1, 3, 5) | 2.25 | (1, 2, 3, 5) | 2.75 |
| (1, 2) | 1.5 | (3, 4) | 3.5 | (0, 1, 4, 5) | 2.50 | (1, 2, 4, 5) | 3.00 |
| (1, 3) | 2.0 | (3, 5) | 4.0 | (0, 2, 3, 4) | 2.25 | (1, 3, 4, 5) | 3.25 |
| | | (4, 5) | 4.5 | | | (2, 3, 4, 5) | 3.50 |

Frequency distribution of $P_1$

| Value | Frequency |
|---|---|
| 0 | 1 |
| 1 | 1 |
| 2 | 1 |
| 3 | 1 |
| 4 | 1 |
| 5 | 1 |

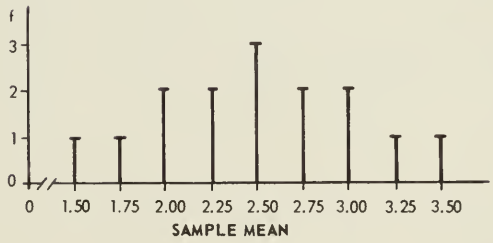
Frequency distribution of $P_2$

| Sample mean | Frequency |
|---|---|
| 0.5 | 1 |
| 1.0 | 1 |
| 1.5 | 2 |
| 2.0 | 2 |
| 2.5 | 3 |
| 3.0 | 2 |
| 3.5 | 2 |
| 4.0 | 1 |
| 4.5 | 1 |




Frequency distribution of $P_4$

| Sample mean | Frequency |
|---|---|
| 1.50 | 1 |
| 1.75 | 1 |
| 2.00 | 2 |
| 2.25 | 2 |
| 2.50 | 3 |
| 2.75 | 2 |
| 3.00 | 2 |
| 3.25 | 1 |
| 3.50 | 1 |




The conclusions to be drawn from this illustration are:

(i) As the sample size increases, the frequency distribution of the population of sample means approaches a normal distribution.

(ii) As the sample size increases, the variability in the population of sample means diminishes.

## INFERENTIAL CONCEPTS

Two inferential tools of great use in statistics are Tests of Hypothesis and Confidence Intervals. Although closely related in their theoretical basis, they serve different needs from an applied viewpoint.

Example:

Suppose a scientist is investigating the number of bacteria per gram in frozen eggs. Suppose also that his main interest is in the mean number of bacteria per gram (call it $\mu$). Two situations can be visualized:

(i) The scientist may have some theory of his own or he may have other facts (another scientist's experiments) that lead him to believe that $\mu = 150$. He may decide to use his observations to test if, in fact, $\mu = 150$ is consistent with his experimental results.

(ii) Or the scientist may want to estimate the mean number from his data and may not be concerned with checking any particular theory. Situation (i) leads to Statistical Tests of Hypothesis and situation (ii) leads to Confidence Intervals.

## PROBABILITY

There are several interpretations of probability. Here we give the approach referred to as *frequentist.* In a long series of throws with a coin, heads and tails will occur approximately equally often. We then say that the probability of a head (or tail) turning up is ½. In a long series of throws with a fair die, each of the six sides will occur in approximately 1/6 of the total number of throws and the probability for any of the numbers is 1/6. Generally the concept of probability can be formulated by saying that in a very long series of trials any event will tend to occur with a relative frequency that is approximately equal to the probability of the event.

## TESTS OF HYPOTHESIS

An example borrowed from criminal legal procedure may help to introduce some concepts.

Null hypothesis: $H_o$: defendant is innocent.

Alternative hypothesis: $H_a$: defendant is guilty.

The jury observes the evidence and reaches a verdict. The two types of error are:

Type I error: an innocent person is found guilty.

Type II error: a criminal is acquitted.

The type I error consists of rejecting the null hypothesis when it is true. The type II error consists of accepting the null hypothesis when the alternative is true.

An example of a statistical test of hypothesis is the following: A coin is known to be either fair (ht) or two headed (hh). After a single toss of the coin, a test of the hypothesis must be set up:

$H_o$: the coin is fair

versus $H_a$: the coin is two headed.

The test of a statistical hypothesis consists of a decision rule to accept or reject the null hypothesis ($H_o$) on the basis of relevant statistical information (outcome) of a single toss of the coin. There are many decision rules that can be used to construct a test. Some are better (more reasonable) than others. Here are three different decision rules, each of which can be used as a statistical test for the null hypothesis that the coin is ht versus the alternative that the coin is hh.

$D_1$: Toss the coin and if t shows up accept $H_o$; if h shows up reject $H_o$.

$D_2$: Toss the coin and if t shows up accept $H_o$; if h shows up accept $H_o$.

$D_3$: Toss the coin and if it rains outside accept $H_o$; if it does not rain outside reject $H_o$.

If we must choose one of the decision rules as a basis for our test, $D_3$, although it is a decision rule, can be disregarded on common sense grounds, since it does not use experimental evidence. However, when $D_1$ and $D_2$ are compared, it is not clear which one is better. Therefore, a measure of the goodness of the decision rule that is to be used as a basis for a statistical test of hypothesis is needed.

Associated with a decision rule are two quantities: $\alpha$ = P (making a type I error) which reads "probability of making a Type I error" and $\beta$ = P (making a type II error). Clearly, a desirable property for a statistical test is that both $\alpha$ and $\beta$ be as low as possible. Thus, the pair ($\alpha$, $\beta$) provides a basis to measure the goodness of a statistical test of hypothesis. Unfortunately, the size of both $\alpha$ and $\beta$ is not enough by itself to select the best one of several available tests of hypothesis. This point will be illustrated by computing $\alpha$ and $\beta$ for decision rules $D_1$ and $D_2$ in the coin example.

For $D_1$,

$\alpha$ = P (making a type I error)
= P (rejecting $H_o$ when it is true)
= P (h shows up with a fair coin)
= $\frac{1}{2}$

$\beta$ = P (making a type II error)
= P (accepting $H_o$ when $H_a$ is true)
= P (t shows up with a two-headed coin)
= 0

Therefore for $D_1$, $\alpha = \frac{1}{2}$ and $\beta = 0$

For $D_2$,

$\alpha$ = P (making a type I error)
= P (rejecting $H_o$ when it is true)
= 0, because according to $D_2$ we never reject $H_o$

$\beta$ = P (making a type II error)
= P (accepting $H_o$ when $H_a$ is true)
= P (heads or tails show up when coin is 2h)
= 1, because heads will always appear, so the event (heads or tails) will always occur.

Therefore:

| | $\alpha$ | $\beta$ |
|---|---|---|
| $D_1$ | $\frac{1}{2}$ | 0 |
| $D_2$ | 0 | 1 |

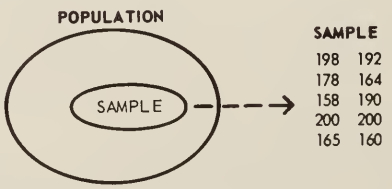
The above table shows that the decision rule ($D_1$ or $D_2$) that should be used cannot be chosen only by considering the probabilities $\alpha$ and $\beta$ that correspond to each decision rule because their relative importance is unknown. Consideration of the economic importance of the type I and type II errors is necessary when choosing which decision rule to use for a particular situation. Further discussion of this point is beyond the scope of these notes.

The probability of making a type I error is called the level of significance of the test, while the quantity, 1 minus P (type II error), is referred to as the power of the test. In statistical methodology, if no economic criteria are available to weigh the importance of the type I and type II errors, the usual way to select a test is to fix the level of significance arbitrarily, say 0.05, to consider only the decision rules for which the level of significance is 0.05, and then to choose the decision rule for which the power is maximum.

## CONFIDENCE INTERVALS

Research workers often have to estimate parameters. An estimate of a parameter given by the corresponding statistic is unlikely to be precisely equal to the parameter, so it is necessary to show the margin of variability to which it is subject. A way to do this is to specify an interval within which we may be "confident" that the parameter lies. Such an interval is called a confidence interval.

Example: A sample of 10 adult Canadians is taken to estimate the mean Canadian adult weight ($\mu$).



If it is assumed that Canadian adult weights are normally distributed, by using a technique available in any textbook on statistics, the 95% confidence interval for $\mu$ can be computed from the sample and is 160.75, 190.25. The meaning of a confidence interval is often misunderstood. It does not mean that the probability is 0.95 that the population mean ($\mu$) lies between 160.75 and 190.25. It means that if many random samples of size 10 are taken from the population of adult Canadian weights and for each sample a 95% confidence interval for the mean is computed according to the above technique, then, in a very long series 95% of the confidence intervals so computed will contain the true mean.

## REFERENCES


1. Amerine, M. A., and C. S. Ough. 1964. The Sensory Evaluation of Californian Wines. Lab. Pract. Vol. 13, No. 8.
2. Baker, R. A. 1964. Taste and Odour in Water. Lab. Pract. Vol. 13, No. 8.
3. Bengtsson, K., and E. Helm. 1953. Principles of Taste Testing. Wallerstein Laboratories Communications.
4. Caul, J. F. 1957. The Profile Method of Flavor Analysis. Adv. in Food Research, 7: 1.
5. Cochran, W.G., and G.M. Cox. 1957. Experimental Designs. John Wiley and Sons, New York, N.Y.
6. Dawson, E. H., J.L. Brogdon, and S. McManus. 1963. Sensory Testing of Differences in Taste. Food Technol. Vol. 17, No. 9.
7. Duncan, D. B. 1955. Multiple Range and Multiple F Tests. Biometrics, Vol. II.
8. Eindhoven, J., D. Peryam, F. Heiligman, and G. A. Baker. 1964. Effect of Sample Sequence on Food Preference. J. Food Sci. Vol. 29, No. 4.
9. Fisher, R. A., and F. Yates. 1942. Statistical Tables. Oliver and Boyd Ltd., Edinburgh and London.
10. Kramer, A., J. Cooler, M. Modery, and B.A. Twigg. 1963. Numbers of Tasters Required to Determine Consumer Preference for Fruit Drinks. Food Technol. Vol. 17, No. 3.
11. Lowe, B. 1963. Experimental Cookery, John Wiley and Sons. New York, N.Y.
12. Pangborn, R. M. V. 1964. Sensory Evaluation of Food at the University of California. Lab. Pract. Vol. 13, No. 7.

13. Pettit, L. A. 1958. Informational Bias in Flavor Preference Testing. Food Technol. Vol. 12, No. 1.

14. Peryam, D. R. 1964. Sensory Testing at the Quartermaster Food and Container Institute. Lab. Pract. Vol. 13, No. 7.

15. Read, D. R. 1964. A Quantitative Approach to the Comparative Assessment of Taste Quality in the Confectionery Industry. Biometrics, 20: 143–155.

16. Robinson, P. 1959. Tests of Significance. Bulletin of Statistical Research Service, Canada Department of Agriculture.

17. Sather, L. A. 1958. Laboratory Flavor Panels. Oregon Agricultural Experiment Station Paper No. 56.

18. Sather, L. A., and L. D. Calvin. 1960. The Effect of Number of Judgments in a Test on Flavor Evaluations for Preference. Food Technol. Vol. 14.

19. Sather, L. A., L. D. Calvin, and A. Tomsma. 1963. Relation of Preference Panels and Trained Panel Scores on Dry Whole Milk. J. Dairy Sci. 46.

20. Scheffé, H. 1952. An Analysis of Variance for Paired Comparisons. J. Am. Statist. Ass. 47: 381–400.

21. Sjostrom, L. B., S. E. Cairncross, and J. F. Caul. 1957. Methodology of the Flavor Profile. Food Technol. Vol. 11, p. 20.

22. Snedecor, G. W. 1956. Statistical Methods. Iowa State College Press, Iowa State College, Ames.

23. Tilgner, D. J. 1962. Anchored Sensory Evaluation Tests – A Status Report. Food Technol. Vol. 16, No. 3.

24. Tilgner, D. J. 1962. Dilution Tests for Odor and Flavor Analysis. Food Technol. Vol. 16, No. 2.

## ADDITIONAL SOURCES OF INFORMATION

Bartlett, F. 1964. The Evaluation of Sensory Experiences. Lab. Pract. Vol. 13, No. 7.

Christie, E. M. 1964. Taste Testing in the C.S.I.R.O. Lab. Pract. Vol. 13, No. 7.

Committee on Sensory Evaluation of the Institute of Food Technology. 1964. Sensory Testing Guide for Panel Evaluation of Food and Beverages.

Ellis, B. H. 1961. A Guide Book for Sensory Testing. Continental Can. Co. Inc.

Gridgeman, N. T. 1959. Pair Comparisons, With and Without Ties. Biometrics, 15: 382–388.

Gridgeman, N. T. 1959. Sensory Item Sorting. Biometrics, 15: 298–306.

Gridgeman, N. T. 1961. A Comparison of Some Taste Test Methods. J. Food Sci. 26: 171–177.

Harper, R. 1964. The Sensory Evaluation of Food and Drink. An Overview. Lab. Pract. Vol. 13, No. 7.

Harries, J. M. 1964. Sensory Testing at the Ministry of Agriculture, Fisheries and Food. Lab. Pract. Vol. 13, No. 7.

Kramer, A., and B. Twigg. 1962. Fundamentals of Quality Control in the Food Industry. The AVI Publishing Co. Inc., Westport, Connecticut.

McCowen, P. 1964. Sensory Testing at Lyons Ltd. Lab. Pract. Vol. 13, No. 8.

Sullivan, F., and J.F. Caul. 1964. Applications of Flavor Profile to Food and Beverage Packaging Problems. Lab. Pract. Vol. 13, No. 7.

Tilgner, D. J. 1964. Sensory Analysis at the Politichnika Gdanska. Lab. Pract. Vol. 13, No. 7.

Wallis, W. A., and H. V. Roberts. 1956. Statistics, New Approach. The Free Press, Glencoe, Chicago, Illinois.

CAL/BCA OTTAWA K1A 0C5
3 9073 00185325 0

INFORMATION
Edifice Sir John Carling Building
930 Carling Avenue
Ottawa, Ontario
K1A 0C7

Canada Post Postage paid
Postes Canada Port payé
Third class
Troisième classe
K1A 0C5
Ottawa

IF UNDELIVERED, RETURN TO SENDER
EN CAS DE NON-LIVRAISON, RETOURNER À L'EXPÉDITEUR